

Linear reduced order method for design-space dimensionality reduction and flow-field learning in hull form optimization

Xinwang Liu^a, Weiwen Zhao^a, Decheng Wan^{a,b,*}

^a Computational Marine Hydrodynamics Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

^b Ocean College, Zhejiang University, Zhoushan, 316021, China

ARTICLE INFO

Keywords:

Dimensionality reduction
Hull form optimization
Proper orthogonal decomposition
Design space
Multi-physics field learning
Sensitivity analysis

ABSTRACT

In the earlier stage of hull form optimization design, a series of design variables is usually needed to control the hull shape, so as to find optimal hull forms with better performance. In the surrogate-based hydrodynamic performance optimization for ships, with the increase of the dimensionality of design space, the number of new sample hulls to construct surrogate model needs to be larger, which will bring a large amount of calculation. Through reduced order method, the dimensionality of the optimization design space can be reduced while keeping the deformation range of the original design space to a great extent, for instance, using the linear combination of a smaller number of bases to represent the deformation range. In addition, in the later stage of hull form optimization design, flow field results of the new sample hulls can be fully utilized to do the dimensionality reduction multi-physics field learning. In this paper, the principle of the Proper Orthogonal Decomposition method is used and briefly introduced, the steps of dimensionality reduction of the design space are shown then, and some important problems for the design-space dimensionality reduction in the specific field of hull form optimization, such as retainability of fixed control points, irrelevance of the relative order of data to dimensionality reduction results, and decision of the new design space range after dimensionality reduction, are deep analyzed. Furthermore, taking the resistance optimization of the modified Wigley ship as an example, the specific application and error analysis of the dimensionality reduction method for design-space dimensionality reduction in the earlier stage of hull form optimization and the multi-physics field learning in the later stage of hull form optimization are given, and the applicability and reliability of the method are demonstrated by analyzing the influence of mode order and sample number on reconstruction effect of the hull shape or flow field, and the prediction effect of flow field for not-in-the-database new hull form in detail. Results show that the linear dimensionality reduction method can reduce samples needed for optimization, thus reduce the amount of calculation for the surrogate-based hull form optimization, and be used for quick prediction of multi-physics fields of any new form in the design space. Furthermore, it can not only be applied to the sensitivity analysis or a Pareto frontier selection in comprehensive performance optimization of hull form based on CFD, but also be implemented in the real-time forecast of the flow field and influence analysis of the ship performance when adjusting the hull form (or hull appendages).

1. Introduction

For the hull form optimization problem, if the number of design variables that control the hull form deformation is small, the hull shape change may be small in a certain design space, that is, the geometry variation is not enough. Therefore, the improvement of the performance index of the optimal hull is relatively limited. On the contrary, if there

are more parameters controlling the change of hull shape, the possible changes of hull form will be more abundant, that is to say, in the high-dimensional design space, more diverse geometry variations will be generated, and the optimal hull with much better performance can be found theoretically.

In order to avoid the difficulty of solving the optimization problems in the high-dimensional design space caused by the curse of

* Corresponding author. Computational Marine Hydrodynamics Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.

E-mail address: dcwan@sjtu.edu.cn (D. Wan).

<https://doi.org/10.1016/j.oceaneng.2021.109680>

Received 10 April 2021; Received in revised form 12 July 2021; Accepted 11 August 2021

0029-8018/© 2021 Elsevier Ltd. All rights reserved.

dimensionality, different kinds of dimensionality reduction technologies for the design space have been proposed and widely applied, including online dimensionality reduction technology and off-line dimensionality reduction technology. The so-called “dimensionality reduction” means to reduce the data dimension as far as possible on the premise that it basically includes all the information involved in the set of high-dimensional data vectors, that is, to basically represent all possible changes in the higher-dimensional design space using a lower-dimensional design space. Through the dimensionality reduction of the design space of hull form deformation parameters (design variables), a smaller number of design variables in the lower-dimensional design space can represent any possible hull shape change without losing much geometry variation information in original high-dimensional design space. Therefore, the design of experiment can be implemented in the new lower-dimensional design space, where the new design variables can basically represent all possible changes in the higher-dimensional design space. For the surrogate-based optimization, the number of new sample hulls needed to evaluate will be reduced due to the reduction of design-space dimensionality, so the calculation cost will be reduced, and then the optimization efficiency will be improved.

Online design space dimensionality reduction technology needs to evaluate the objective function (such as hydrodynamic performance index) or its gradient. [Lukaczyk et al. \(2014\)](#) proposed a search strategy for active dimensionality-reduction subspace. However, such methods could not evaluate and reduce dimensionality of the design space for various hull form deformation methods in advance, that is, they cannot consider the strong relationship that may exist in the hull shape changes controlled only by different design variables.

Off-line design space dimensionality reduction technology mainly considers the variability of the design space, which does not need the evaluation of the objective functions, let alone the gradient information of them.

In terms of the decision of the new feature (independent variable) in the new dimensionality-reduced design space, the off-line dimensionality reduction methods can be divided into feature selection and feature extraction methods. Feature selection is just a filtering of the original features, that is, a subset of independent variables in the original high-dimensional space is chosen, which usually have the greatest influence on the dependent variables. Feature extraction is to condense and construct a new set of features on the basis of original features, that is, it converts the old set of high-dimensional variables into a new set of low-dimensional variables by data analysis. In this paper, the dimensionality reduction method of feature extraction is further elaborated and studied.

Proper Orthogonal Decomposition (POD), also known as Principal Component Analysis (PCA) or Karhunen-Loève Expansion (KLE) ([Pearson, 1901](#); [Wold et al., 1987](#); [Jolliffe, 1986](#); [Diez et al., 2015](#)) is one of the most classic, but still most widely used linear reduced order methods. The basic idea of this method is to construct a lower-dimensional linear space, making the projection of a certain number of high-dimensional data set on it as large as possible. In other words, mapping the high-dimensional data set into a low-dimensional data set by constructing a transformation, while making the low-dimensional data set keep the change (variance) information of the original high-dimensional data set as much as possible.

However, when there are complex nonlinear relationships between design variables, some linear dimensionality reduction methods may not work well. In order to reduce the dimensionality of the design space under the premise of capturing the nonlinear information in the original high-dimensional design space, the nonlinear dimensionality reduction methods were proposed and developed successively after the linear method, such as Local PCA (LPCA) ([Lloyd, 1982](#); [Kambhatla et al., 1997](#); [Luxburg, 2004](#)), Kernel PCA (KPCA) ([Scholkopf, 1998](#)), and Deep auto-encoder (DAE) ([Hinton et al., 2006](#)).

Both linear and nonlinear dimensionality reduction methods have their advantages and disadvantages. The main advantage of the nonlinear method is that it can capture the nonlinear relationship

between the original high-dimensional data while reducing the dimensionality. However, it may need some additional parameter settings and debugging in the calculation to ensure that the relative error before and after dimensionality reduction is small enough, so the calculation stability is not as good as the linear method. In addition, the basis in the linear method usually has a clear physical concept, and the new samples can be reconstructed through its linear superposition conveniently, such as the changes of hull geometry, while the nonlinear mapping relationship in the nonlinear method is only on the mathematical level, and its physical significance is not obvious.

Furthermore, both linear and nonlinear dimensionality reduction methods have been widely used in image compression, signal processing, text classification and many other engineering fields. However, the application of dimensionality reduction technology in the field of hull form optimization is relatively rare.

[Chen et al. \(2014\)](#) optimized the Delft catamaran calm-water drag under a certain speed. By using KLE method on two design spaces of 20 Free Form Deformation (FFD) parameters, the dimensionalities of the new design spaces after dimensionality reduction become 4 and 6, respectively, and the optimal hulls in the two new design spaces have a 9.63% and 6.89% reduction respectively.

[Diez et al. \(2015\)](#) briefly gave the basic principle of design-space dimensionality reduction and an optimization application of calm-water drag for the high-speed catamaran. By using KLE method on the 20-dimensional design space, the dimensionalities of the new design spaces after dimensionality reduction are controlled no more than 4. Optimization results show that there are significantly positive correlations between the total drag reduction degree and dimensionality of the new design space.

[D’Agostino et al. \(2020\)](#) used KLE method to optimize the calm-water drag of DTMB-5415 ship by reducing the dimensionality of the design spaces composed of design variables of FFD, Radial Basis Function (RBF) and Global Modification Function (GMF) methods. Under the precondition of keeping 95% geometric changes in the original design space, after dimensionality reduction, the final design space dimensionalities are reduced by 86%, 16% and 37%, respectively.

As a matter of fact, the application of dimensionality reduction technology in hull form optimization is more than design-space dimensionality reduction of hull form deformation.

[Serani et al. \(2016, 2018\)](#) gave examples of optimizing the calm-water drag of DTMB-5415 ship by using the potential flow theory for performance evaluation. Considering the change of the hull geometry and physics field, several reduced order methods such as KLE, LPCA, KPCA and DAE were used for multi-physics field learning, with the number of geometric and physical field information grid points 5101. Under the premise that the overall relative error is no more than 5%, the number of bases that contain the multi-physics field information are 18, 14, 15, 15, respectively, by the four dimensionality reduction methods, indicating that there is a strong correlation in the geometrical-physical field.

To sum up, in order to reduce the number of samples to evaluate, and improve the surrogate-based optimization efficiency, in the earlier stage of the hull form optimization, design-space dimensionality reduction can be implemented. After evaluations of the new sample hulls determined by design of experiment, the geometrical-physical field database can be naturally formed and the flow-field dimensionality-reduction learning can be carried out further. Therefore, the application of reduced order method can run through from the earlier stage to the later stage of the hull form optimization process, but there are few relevant studies at present. This paper aims to give specific steps and quantitative error analysis of dimensionality reduction method used in the earlier stage of hull form optimization and the multi-physics field learning in the later stage of hull form optimization. Furthermore, deep discussions on some important problems, such as retainability of fixed control points, irrelevance of the relative order of data to dimensionality reduction results, and decision of the new design space range after

dimensionality reduction, in the specific field of hull form optimization, are given in detail.

2. Basic principles of linear dimensionality reduction method

In the process of hull form optimization, after determining the hull form deformation methods and their corresponding design space, a series of new hull forms can be obtained based on the design of experiment results, that is, a hull form database is obtained. Consider each new hull form as a sample and arrange the three-dimensional grid point coordinates of each hull form in the same order, for example $x_i^T = (x_1^i, y_1^i, z_1^i, x_2^i, y_2^i, z_2^i, \dots, x_{Np}^i, y_{Np}^i, z_{Np}^i)$, where Np is the number of grid points on the hull surface, a series of column vectors can be got. In order to use a linear combination of very few basis vectors to characterize the change of the hull form to the initial hull in the design space, the most important thing is to determine the basis vectors and the number of basis.

Obviously, for all samples in the database, we can obtain the average hull form (geometric field) \bar{x} , and regard the geometric field of any sample hull form as a linear superposition of the average geometric field and an infinite number of unit orthogonal basis modes as follows,

$$x_i = \bar{x} + \sum_{j=1}^{\infty} a_{ij} u_j \quad (1)$$

where

$$a_{ij} = (x_i - \bar{x})^T u_j, \quad (u_i, u_j) = \delta_{ij} \quad (2)$$

If only M (finite) basis modes u_1, u_2, \dots, u_M are selected, the reconstructed geometric field x_i^{rec} can be expressed as

$$x_i^{rec} = \bar{x} + \sum_{j=1}^M a_{ij} u_j \quad (3)$$

Because of the truncation of the number of series items, a ‘‘cutoff error’’ is produced. Considering that the deviation between the real geometric field and the reconstructed geometric field of all samples in the database should be as small as possible, the total error $E_M(u_1, u_2, \dots, u_M)$ is measured by the square sum of the modulus length of the error vector in Euclidean space, namely

$$E_M(u_1, u_2, \dots, u_M) = \sum_{i=1}^N \|x_i - x_i^{rec}\|^2 \quad (4)$$

and we can get

$$\begin{aligned} E_M(u_1, u_2, \dots, u_M) &= \sum_{i=1}^N \left\| x_i - \bar{x} - \sum_{j=1}^M a_{ij} u_j \right\|^2 \\ &= \sum_{i=1}^N \|x_i - \bar{x}\|^2 + \sum_{i=1}^N \left(\sum_{j=1}^M a_{ij} u_j \right)^T \left(\sum_{j=1}^M a_{ij} u_j \right) - 2 \sum_{i=1}^N (x_i - \bar{x})^T \left(\sum_{j=1}^M a_{ij} u_j \right) \end{aligned} \quad (5)$$

It can be seen from the above equation that the total error is divided

into three parts: for a certain sample set, E_1 is a constant. Consider the remaining two parts:

$$\begin{aligned} E_2 &= \sum_{i=1}^N (u_1^T a_{i1}^T + u_2^T a_{i2}^T + \dots + u_M^T a_{iM}^T) (a_{i1} u_1 + a_{i2} u_2 + \dots + a_{iM} u_M) \\ &= \sum_{j=1}^M u_j^T \left(\sum_{i=1}^N a_{ij}^T a_{ij} \right) u_j \\ &= \sum_{j=1}^M u_j^T \left[\sum_{i=1}^N u_j^T (x_i - \bar{x}) (x_i - \bar{x})^T u_j \right] u_j \\ &= \sum_{j=1}^M \left[\sum_{i=1}^N u_j^T (x_i - \bar{x}) (x_i - \bar{x})^T u_j \right] u_j^T u_j \\ &= \sum_{j=1}^M u_j^T \left[\sum_{i=1}^N (x_i - \bar{x}) (x_i - \bar{x})^T \right] u_j \end{aligned} \quad (6)$$

$$\begin{aligned} E_3 &= -2 \sum_{i=1}^N (x_i - \bar{x})^T \left[\sum_{j=1}^M (x_i - \bar{x})^T u_j \right] u_j \\ &= -2 \sum_{i=1}^N \sum_{j=1}^M u_j^T (x_i - \bar{x}) (x_i - \bar{x})^T u_j \\ &= -2 \sum_{j=1}^M u_j^T \left[\sum_{i=1}^N (x_i - \bar{x}) (x_i - \bar{x})^T \right] u_j \end{aligned} \quad (7)$$

Finally, we have

$$\begin{aligned} E_M &= E_1 + E_2 + E_3 \\ &= \sum_{i=1}^N \|x_i - \bar{x}\|^2 - \sum_{j=1}^M u_j^T \left[\sum_{i=1}^N (x_i - \bar{x}) (x_i - \bar{x})^T \right] u_j \end{aligned} \quad (8)$$

Denote the difference between the geometric field of each sample hull form x_i and the average geometry \bar{x} as $\tilde{x}_i = x_i - \bar{x}$, and define the matrix $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N)$, then

$$\tilde{X} \tilde{X}^T = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_N \end{pmatrix}^T = \sum_{i=1}^N \tilde{x}_i \tilde{x}_i^T = \sum_{i=1}^N (x_i - \bar{x}) (x_i - \bar{x})^T \quad (9)$$

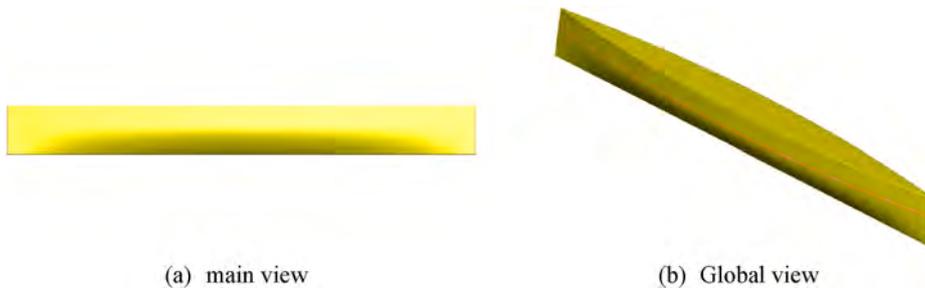
Therefore, the total error can be denoted as

$$E_M = \sum_{i=1}^N \|x_i - \bar{x}\|^2 - \sum_{j=1}^M u_j^T \tilde{X} \tilde{X}^T u_j \quad (10)$$

Assuming that M is certain, the constrained optimization problem on the basis u_1, u_2, \dots, u_M can be defined as follows:

$$\begin{aligned} \min E_M(u_1, u_2, \dots, u_M) &= \sum_{i=1}^N \|x_i - \bar{x}\|^2 - \sum_{j=1}^M u_j^T \tilde{X} \tilde{X}^T u_j \\ &\text{s.t. } u_j^T u_j = 1 \end{aligned} \quad (11)$$

Using the Lagrange multiplier method, the Lagrange function corresponding to the above objective function is defined as



(a) main view

(b) Global view

Fig. 1. Geometry model of Wigley.

$$L(u_1, u_2, \dots, u_M; \lambda_1, \lambda_2, \dots, \lambda_M) = E_M - \sum_{j=1}^M \lambda_j (1 - u_j^T u_j) \quad (12)$$

Let the partial derivatives of Lagrange function with respect to each basis be 0, i.e.

$$\frac{\partial L}{\partial u_j} = 0 \quad (13)$$

we have

$$\tilde{X}\tilde{X}^T u_j = \lambda_j u_j \quad (14)$$

In other words, the basis we're looking for should be the eigenvectors corresponding to the eigenvalues of matrix $S = \tilde{X}\tilde{X}^T \in \mathbb{R}^{3Np \times 3Np}$. Generally, there are two ways to determine λ_j . One is to directly calculate the eigenvalue of square matrix S , whose dimension is the number of grid point coordinates on the hull surface; the other is to solve the singular value of matrix \tilde{X} (or \tilde{X}^T) according to the relation between the singular value of \tilde{X} (or \tilde{X}^T) and the eigenvalue of S . From the perspective of improving the efficiency of numerical calculation, since the number of grid point coordinates stored in geometric field is generally greater (even far greater) than the number of samples in the database, in order to calculate the eigenvalue of matrix S , we can solve the singular value decomposition of matrix \tilde{X} instead, namely:

$$\tilde{X} = UDV^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T \quad (15)$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, then we have

$$\begin{aligned} \tilde{X}\tilde{X}^T u_j &= (UDV^T VD^T U^T) u_j \\ &= (UD^2 U^T) u_j \\ &= U \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0) U^T u_j \\ &= \sigma_j^2 u_j \end{aligned} \quad (16)$$

where the basis u_j happens to be the j -th column of the matrix U obtained by singular value decomposition of \tilde{X} .

Next, the number of intercepted basis modes M should be determined. From the perspective of capturing the total energy (information) of the geometric field, the energy contained in the basis u_j can be defined as the sum of squares of the u_j -direction projection of the geometric field vector corresponding to each sample:

$$\begin{aligned} \sum_{i=1}^N \left\| \left(\tilde{x}_i, u_k \right) \right\|^2 &= \sum_{i=1}^N \left\| \left(\sum_{j=1}^{\infty} a_{ij} u_j, u_k \right) \right\|^2 \\ &= \sum_{i=1}^N a_{ik}^2 \\ &= \sum_{i=1}^N \left(\tilde{x}_i^T u_k \right)^T \left(\tilde{x}_i^T u_k \right) \\ &= u_k^T \tilde{X}\tilde{X}^T u_k \\ &= \sigma_k^2 \end{aligned} \quad (17)$$

According to the above formula, we can see that the energy contained in the basis u_j is exactly the eigenvalue corresponding to the basis u_j as the eigenvector of matrix S . Therefore, define the percentage of energy truncation can be defined as

$$E_n = \frac{\sum_{j=1}^M \sigma_j^2}{\sum_{j=1}^{3Np} \sigma_j^2} \times 100\% \quad (18)$$

where M can be regarded as the smallest positive integer that meets the inequality $E_n > 95\%$. At this point, we believe that the eigenvectors corresponding to the first M maximum eigenvalues of matrix S contain 95% of the geometric field information in the whole database.

So far, we almost get the reconstructed geometric field after dimensionality reduction. In fact, the derivation of KLE given by Diez et al. (2015) is based on setting up the energy integral as a functional, and uses variational method to obtain the conditions that the basis mode vectors should satisfy. In this paper, the so-called POD or PCA method is used directly to minimize the overall reconstruction error of the hull geometry database, and an optimization problem with constraints can be established, and conditions that the basis mode vectors should meet can be got finally. It turns out that they are essentially the same. Therefore, we can say KLE, PCA, and POD are just called differently in different fields, but essentially the same method.

3. Further discussion on dimensionality reduction of hull form optimization design space

In the following, several important problems arising from the dimensionality reduction of hull form optimization design space in the early stage are further discussed.

3.1. Retainability of fixed control points

When the hull is deformed, it is often required that the grid points at some positions cannot be moved, such as the design waterline, middle longitudinal section and some transverse sections. Such special positions can be limited by using the traditional hull form deformation methods to ensure that the grid points do not move at these positions. However, after the dimensionality reduction of the design space, if the new hull form controlled by the new design variables has some slight deviation in the grid points' coordinates of special positions, it will make the deformed hull do not meet the design requirements.

If the coordinate in a certain direction of one grid point needs to remain unchanged, the coordinate of this position (such as the k -th coordinate, $1 \leq k \leq 3Np$) in each sample hull form vector x_i obtained by the original hull form deformation methods remains unchanged, then the coordinate of this position in the average new hull form vector \bar{x} is also the original value, and the columns of the matrix $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N)$ are defined by $\tilde{x}_i = x_i - \bar{x}$.

So, we have

$$\tilde{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,N} \\ x_{21} & x_{22} & \dots & x_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{k,1} & x_{k,2} & \dots & x_{k,N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{3Np,1} & x_{3Np,2} & \dots & x_{3Np,N} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,N} \\ x_{21} & x_{22} & \dots & x_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{3Np,1} & x_{3Np,2} & \dots & x_{3Np,N} \end{bmatrix} \quad (19)$$

then

$$\tilde{X}\tilde{X}^T = \begin{bmatrix} * & * & * & * \\ 0 & 0 & \dots & 0 \\ * & * & * & * \end{bmatrix} \begin{bmatrix} * & 0 & * \\ * & 0 & * \\ * & \vdots & * \\ * & 0 & * \end{bmatrix} \quad (20)$$

Assume α be the eigenvector of the matrix $\tilde{X}\tilde{X}^T$ with respect to eigenvalue λ , then

$$\begin{bmatrix} * & * & * & * \\ 0 & 0 & \dots & 0 \\ * & * & * & * \end{bmatrix} \begin{bmatrix} * & 0 & * \\ * & 0 & * \\ * & \vdots & * \\ * & 0 & * \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \vdots \\ \alpha_{3Np} \end{pmatrix} = \lambda \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \vdots \\ \alpha_{3Np} \end{pmatrix} \quad (21)$$

Considering that the eigenvalues corresponding to each basis mode obtained by dimensionality reduction are all positive real numbers, it is only necessary to consider $\lambda > 0$, then

$$\begin{pmatrix} * \\ * \\ \vdots \\ 0 \\ \vdots \\ * \end{pmatrix} = \lambda \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \vdots \\ \alpha_{3Np} \end{pmatrix} \quad (22)$$

Therefore, we get

$$\alpha_k = 0 \quad (23)$$

That is, the k -th coordinate of the corresponding basis vector is 0. For arbitrary reconstructed new hull form x_i^{ec} after dimensionality reduction, according to Eq. (2), it is easy to know that the k -th coordinate is still the original value. Therefore, for any reconstructed new hull form, the fixed points for the primitive hull form deformation methods keep unchanged after dimensionality reduction.

3.2. Irrelevance of the relative order of data to dimensionality reduction results

When the hull is deformed, the coordinates of the grid points on the hull surface will change to some extent, but the topological relationship among the grid points will not change. However, in actual dimensionality reduction operation, there is no clear rule for the sequence of each new hull form in the database when the coordinates are written into column vector, and there is also no clear rule for the arrangement of different new hull forms, that is, the column vectors. As a result, it is necessary to discuss whether the uncertainty of relative position will have an impact on the new reconstructed hull forms obtained by dimensionality reduction of the design space.

Of course, the first thing we must ensure is that the orders of each new coordinate point for each new hull form and original hull form are exactly the same. Under this premise, the so-called ‘‘uncertainty of relative position’’ is actually the uncertainty of the relative position of the rows or columns in the matrix $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N)$.

Without loss of generality, the following analysis will only take the uncertainty of the relative position of the rows as an example.

Assuming that the deviation matrices obtained are \tilde{X} and \tilde{X}' respectively before and after exchanging the locations of the coordinates of each new hull form, then it's obviously that \tilde{X}' is the row transformation of \tilde{X} , which is the result of several elementary row transformations that only exchange certain two rows, and it can be expressed in matrix form as

$$\tilde{X}' = H_n H_{n-1} \dots H_2 H_1 \tilde{X} = H \tilde{X} \quad (24)$$

where $H_k^{-1} = H_k^T, k = 1, 2, \dots, n$.

Then we have

$$H^{-1} = H^T \quad (25)$$

Let $S' = \tilde{X}' \tilde{X}'^T$, then

$$\begin{aligned} S' &= \tilde{X}' \tilde{X}'^T \\ &= H \tilde{X} \tilde{X}^T H^T \\ &= H \tilde{X} \tilde{X}^T H^{-1} \\ &= H S H^{-1} \end{aligned} \quad (26)$$

Therefore $S' \sim S$, and the eigenvalues of the two matrices are the same. Assume α be the eigenvector of the matrix S with respect to eigenvalue λ , i.e.

$$S\alpha = \lambda\alpha \quad (27)$$

then

$$S' H \alpha = H S \alpha = \lambda H \alpha \quad (28)$$

That is to say, $H\alpha$ is the eigenvector of the matrix S' with respect to eigenvalue λ , and $H\alpha$ is just the basis mode after exchanging locations of the coordinates of α . Therefore, we can know that after exchanging the locations of the coordinates of each new hull form, the order of reconstructed hull form vector by design-space dimensionality reduction has the same change, but they essentially both represent the same new hull form. Consequently, the relative order of data is irrelative to the dimensionality reduction results.

Admittedly, the dimensionality reduction process cannot be directly guided by the proof given above. If not given, however, we can't ensure theoretically that different researchers use the same database (such as the new samples ship hulls and the corresponding geometrical fields) to do the dimensionality reduction operation and have the same dimensionality reduction results under different data arrangements. This proof demonstrates the irrelevance of data order, allowing researchers to set the order of data arbitrarily.

3.3. Decision of the new design space range after dimensionality reduction

After dimensionality reduction of the design space, the dimensionality of the design space reduces, and the design variables are changed from the original hull form deformation parameters to the coefficients of the basis modes. In order to ensure that the deformation in the new design space after dimensionality reduction does not exceed the deformation range controlled by the original design space, the ranges of new design variables after dimensionality reduction need to be further determined. Otherwise, it is very likely to produce unreasonable (un-real) new hull forms.

According to Eqs. (2) and (3), it is easy to know that the approximate range of the basis coefficient is

$$a_j \in \left[\min_{x_i \in X} (x_i - \bar{x})^T u_j, \max_{x_i \in X} (x_i - \bar{x})^T u_j \right], i = 1, 2, \dots, N \quad (29)$$

That is, the value range of each basis coefficient can be estimated through N new sample ship hull forms. In order to ensure the accuracy of range estimation, N new samples in the database should cover the original design space as much as possible. Therefore, the ‘‘dense sampling’’ strategy is adopted in this paper, for example $N = 5000$, and the basis coefficients obtained from 5000 new sample hulls are projected in two dimensions to obtain the spatial distribution of the basis coefficients, and value ranges of the new design variables in the dimensionality-reduction design space can be determined. As a matter of fact, there may be additional constraints on the design variables. Therefore, the new design space may no longer be a regular hypercube design space. If one ignores this, the new design space (hypercube) may include some ‘‘new’’ ship forms, which may be beyond the possible hull deformations controlled by the original design space and may not satisfy some geometric constraints of hull deformation controlled by the original design space, such as the relative change of wet surface area and other constraints, or may not be like practical hull forms, such as having strange bulbous bow shape.

Therefore, a practical method of two-dimensional projections between the new design variables will be proposed in this paper to keep the hull deformation in the new design space within the limits of the original design space as much as possible. Later, in the new design variable with constraints, in order to ensure no computational costs waste, the method of Sequential Constrained Monte-Carlo (Golchi et al., 2016) is used to do the design of experiment in the constrained design space.

4. Application in design-space dimensionality reduction in earlier stage of hull form optimization

In the following, we give an example of using the linear dimensionality reduction method to reduce the dimensionality of the hull form optimization design space to illustrate its applicability in earlier stage of

Table 1
Main particulars of Wigley.

Parameter	Symbol and unit	Value
waterline length	L_{wl} (m)	4
breadth	B (m)	0.4
draught	T (m)	0.25
molded depth	D (m)	0.41
drainage volume	∇ (m ³)	0.177
wet surface area	S (m ²)	2.379

hull form optimization.

4.1. Basic information of the mother ship

Wigley ship is a kind of parabolic mathematical hull form commonly used in the field of ship research, which is regarded as the initial (mother) ship in this paper. The hull is slender, which conforms to the small perturbation assumption of linear potential theory, and has abundant experimental test results to refer. The ship has no specific design speed and can be numerically simulated over a wide range of speed and compared with the experimental results. Its three-dimensional model is shown in Fig. 1, and the main particulars are shown in Table 1.

Since it conforms to small disturbance assumption of the linear theory, when it sails in calm water under high speed, its free surface wave elevation is obvious. At this time, the wave-making drag accounted for the proportion of the total drag is relatively big, and the potential-flow theory, such as Neumann-Michell (NM) potential flow theory (Noblesse et al., 2013), can be used for rapid prediction of wave-making drag. Furthermore, the 1957 ITTC plate frictional drag coefficient formula can be used to get the approximate frictional drag, and the total drag of the ship hull based on the potential flow theory can be got finally, which is the sum of the wave-making drag and the frictional drag.

4.2. Definition of optimization problem

In this section, after generating a basic bulbous bow on the basis of the original Wigley ship (Liu et al., 2019), the total drag of a modified Wigley ship can be further optimized under the speed Froude number 0.3. The Radial Basis Function (RBF) method is used to carry out the hull form deformation at the bulbous bow, and the shifting method is used to adjust the Section Area Curve (SAC) of the whole ship, where the cross-section of the front and aft halves can be independently translated along the direction of the ship length, as shown in Fig. 2.

The Sobol (1979) sampling method is adopted to sample in the 7-dimensional design space. The ranges of design variables are shown in Table 2. All the range values of variables are dimensionless values with respect to the ship model waterline length. Wherein, x corresponds to the coordinate of point P_1 in the direction of the ship length, z corresponds to the coordinate of point P_1 in the direction of ship draught, and y corresponds to the coordinate of point P_2 in the direction of the ship width in Fig. 2. It should be noted that the moveable control points along the ship width are moved symmetrically about the central sheer plane of the hull. Using the RBF deformation method, these deformation parameters control the length, height and width of the bulbous bow respectively.



Fig. 2. Deformation of a modified Wigley based on RBF and shifting methods.

For the shifting method (Yang et al., 2016), modified function g is introduced to modify the SAC of ship hull and new hull forms can be obtained by shifting hull lines of each station along the x direction:

$$g = \begin{cases} \alpha_1 \left[0.5 \left(1 - \cos 2\pi \frac{x - \alpha_2}{\alpha_2 - x_1} \right) \right]^{0.5} & , x_1 \leq x < \alpha_2 \\ -\alpha_1 \left[0.5 \left(1 - \cos 2\pi \frac{x - \alpha_2}{\alpha_2 - x_2} \right) \right]^{0.5} & , \alpha_2 \leq x \leq x_2 \\ 0 & , \text{elsewhere} \end{cases} \quad (30)$$

where x_1, x_2 represents the start and end points of the shifting range in x direction, α_1 represents the amplitude of the modified function, and α_2 represents the fixed point (section) in the shifting range. The SAC of the initial hull and the deformed hull is shown by the black dot line and green solid line in Fig. 3.

For the optimization problem here, there exist two shifting regions, namely fore part (region 1) and aft part (region 2). In shifting region 1, according to Eq. (30), the modified function g_1 can be written by setting $x_1 = 0, x_2 = 0.5$. Likewise, in shifting region 2, according to Eq. (30), the modified function g_2 can be written by setting $x_1 = -0.5, x_2 = 0$.

Here, α_1 and α_3 represent the amplitudes of the modification

Table 2
Design variables and their ranges.

Design variable	Lower bound	Upper bound	Note
x	0.515	0.545	Coordinate of P_1 in x direction
z	-0.049	-0.0344	Coordinate of P_1 in z direction
y	0.005	0.021	Coordinate of P_2 in y direction
α_1	-0.02	0.02	Amplitude of the modification function for the fore half body
α_2	0.2	0.3	Location of fixed cross section for the fore half body
α_3	-0.02	0.02	Amplitude of the modification function for the aft half body
α_4	-0.3	-0.2	Location of fixed cross section for the aft half body

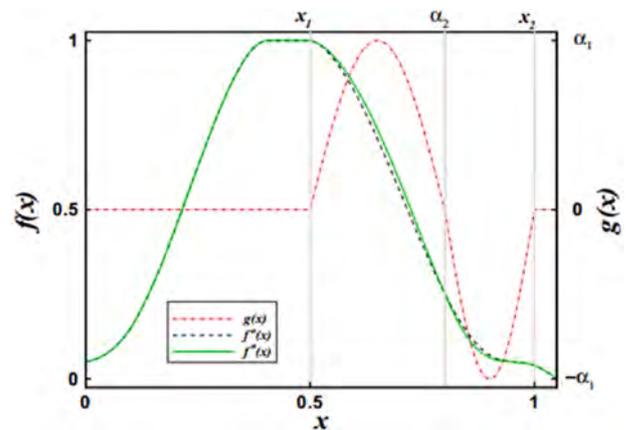


Fig. 3. SAC comparison of the original and deformed hulls.

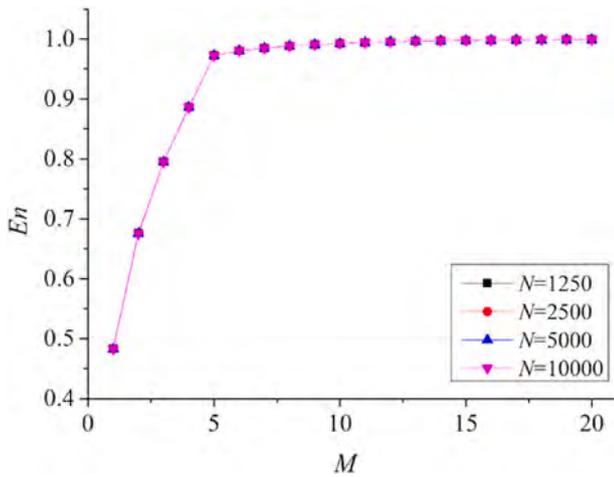


Fig. 4. The variation of En with M with different sample numbers N .

functions for SAC of the fore and aft halves in the shifting method, while α_2 and α_4 represent the locations of the fixed cross sections for the fore and aft halves in shifting method.

The objective function of this optimization problem is the calm-water total drag R_t at the sailing Froude number 0.3, as shown below:

$$\min f_{obj} = R_t \tag{31}$$

4.3. Optimization results and analysis with design-space dimensionality reduction

According to the deformation methods and the design variables, the original design space is determined, and dense samplings by Sobol method can be got, then a large number of new hull forms can be further obtained according to the design variable values of the sample points within the design space, which can build a geometry field database in order to use in the design-space dimensionality reduction process.

Firstly, the influence of different dense sampling numbers on dimensionality reduction of design space should be discussed.

Set $N = 1250, 2500, 5000, 10000$ respectively. After dimensionality reduction by these numbers of new sample hulls, and combining with

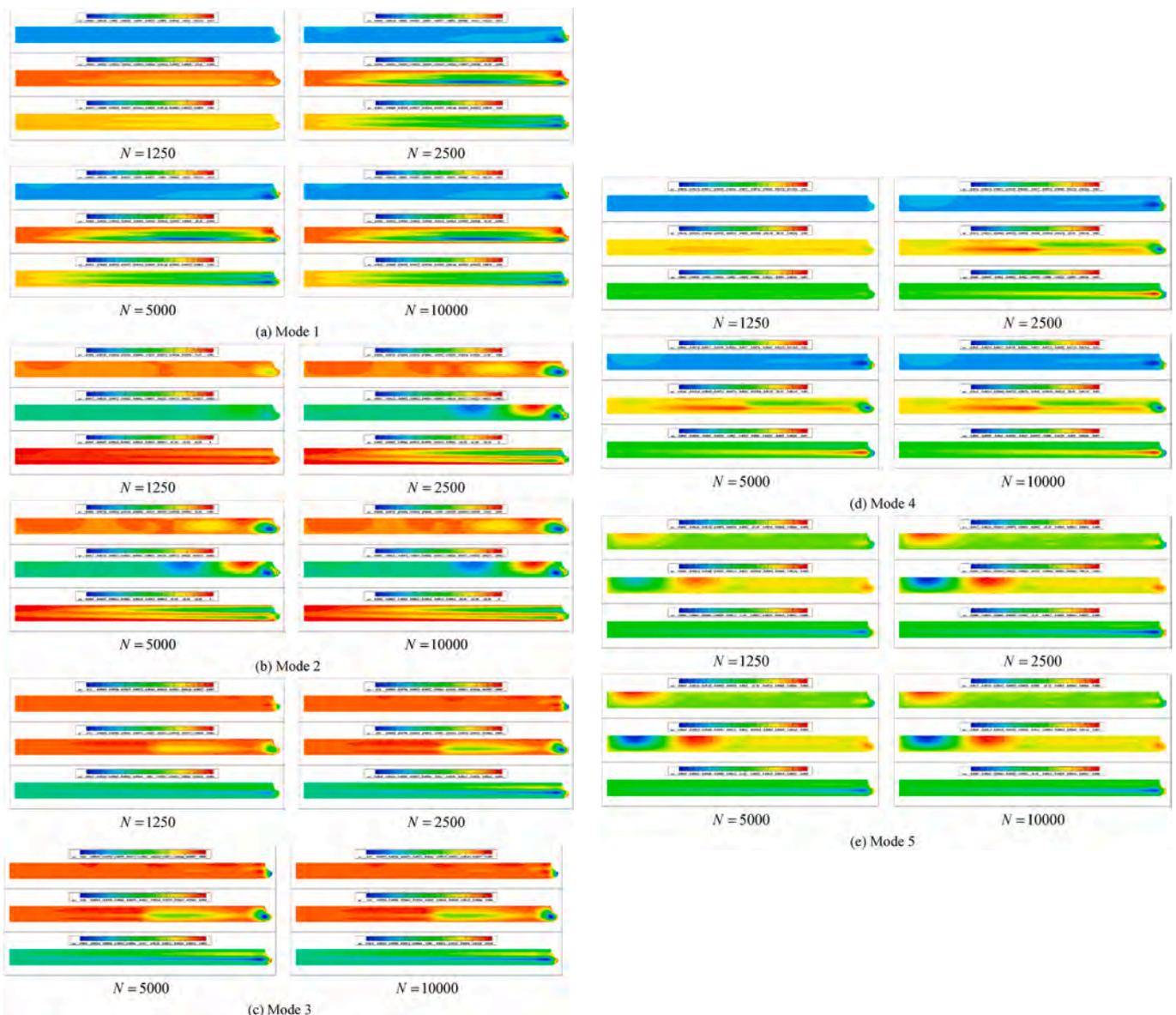


Fig. 5. Basis of each mode obtained by different sample hulls' sum N .

Table 3
New design variables and their ranges after dimensionality reduction.

Design variable	Lower bound	Upper bound	Note
a_1	-0.273	0.177	coefficient of basis u_1
a_2	-0.120	0.111	coefficient of basis u_2
a_3	-0.133	0.086	coefficient of basis u_3
a_4	-0.073	0.073	coefficient of basis u_4
a_5	-0.061	0.061	coefficient of basis u_5

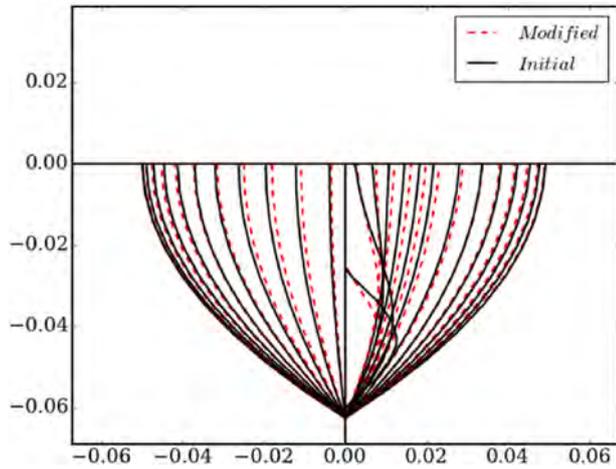


Fig. 6. Comparison of transverse hull lines of initially modified and new deformed hulls.

Eq. (18), the variation of En with the number of bases M and different sample numbers N can be obtained, as shown in Fig. 4.

As can be seen from Fig. 4, for a large sample number N , selecting $M = 5$ can satisfy the energy truncation percentage greater than 95%. That is to say, we can think that the eigenvectors corresponding to the first five maximum eigenvalues of matrix S contain 95% of the information of all possible changes of hull forms in the original design space.

The basis vectors (modes) obtained by dimensionality reduction using the database constructed by different sample points N are shown in Fig. 5. For convenience of display, the components of three directions in the spatial Cartesian coordinate system are given on each grid point of the original modified Wigley ship hull. Since the length (Euclid-norm) of each basis vector is limited to 1, when reflected on the hull surface, the absolute coordinates of each grid point values are quite small. Therefore, according to the maximum of each basis coefficient, each basis vector times the maximum of the corresponding basis coefficient, which can also show the distribution law of the grid point coordinates change (i.e. the hull geometric deformation) corresponding to different basis modes. It is not difficult to find that when N is greater than 2500, the pictures tend to be consistent, indicating that the number of dense sampling N is sufficient.

We select $N = 5000$ as the sample number of the database, and the obtained basis modes are used for further study. Table 3 below gives the approximate value ranges of design variables corresponding to the new design space after dimensionality reduction obtained from Eq. (29).

Taking a deformed hull form in the database as an example, the influence of different mode number M on the hull form reconstructed error is analyzed. As shown in Fig. 6, the solid black line represents an initially modified Wigley ship hull, and the dotted red line represents a deformed ship hull in the database, which is controlled by a total of 7 design variables of RBF and shifting methods.

Since $K = 1, 2, 3, 4, 5$ can be chosen, the contribution of each basis mode to the final reconstruction of the hull form is analyzed below. Set $K = 1, 2, 3, 4, 5$ respectively, and the reconstructed hull form can be

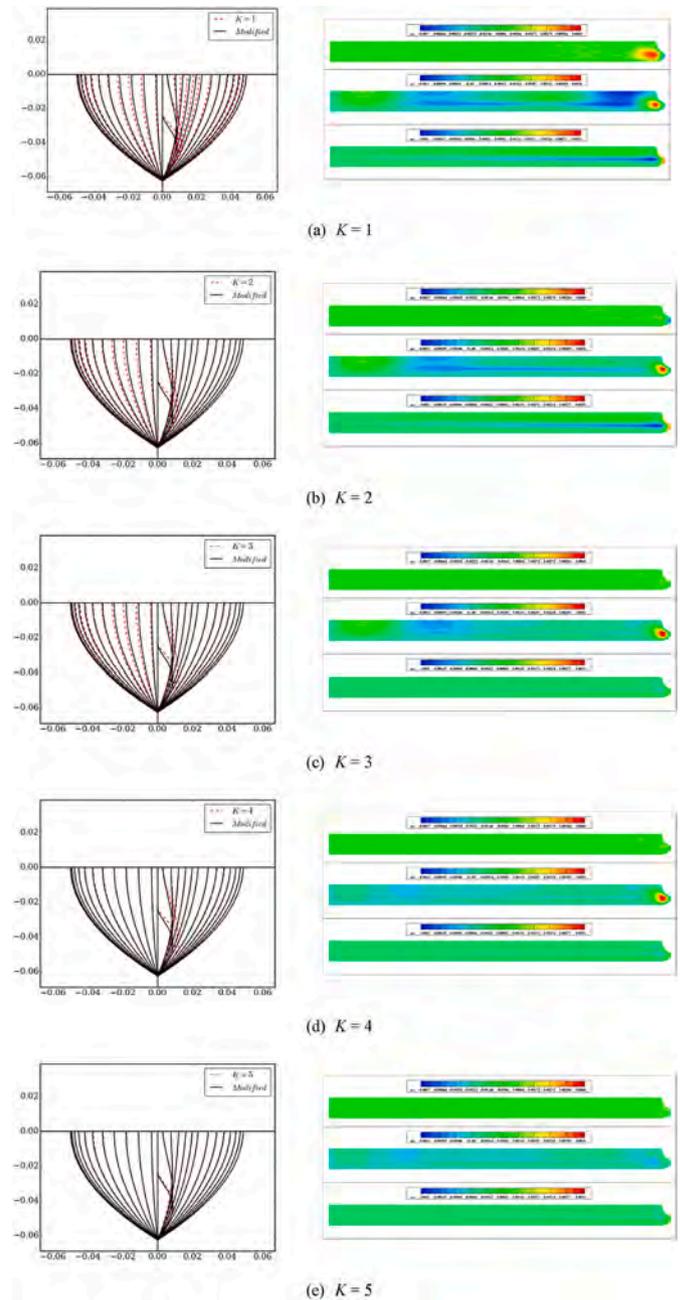


Fig. 7. Hull line comparison (left) and coordinate error (right) of the initial and reconstructed hulls.

obtained by the following equation:

$$x^{rec}(K) = \bar{x} + \sum_{j=1}^K a_j u_j \quad (32)$$

As shown in Fig. 7, the body line comparison and absolute error of the grid point coordinates in three directions (reconstructed deformed hull in red minus target deformed hull in black) of the two hulls are given. Due to the fact that before and after deformation of a hull, the topological relationship of the grid points between the two hulls remains the same, for display convenience, the absolute error of the grid point coordinates in three directions of the two hulls is shown on the grid points of the preliminary modified Wigley hull (with an initially-generated bulbous bow). It can be seen that with the increase of the number of selected basis modes, the absolute error between the real and reconstructed deformed hull decreases gradually.

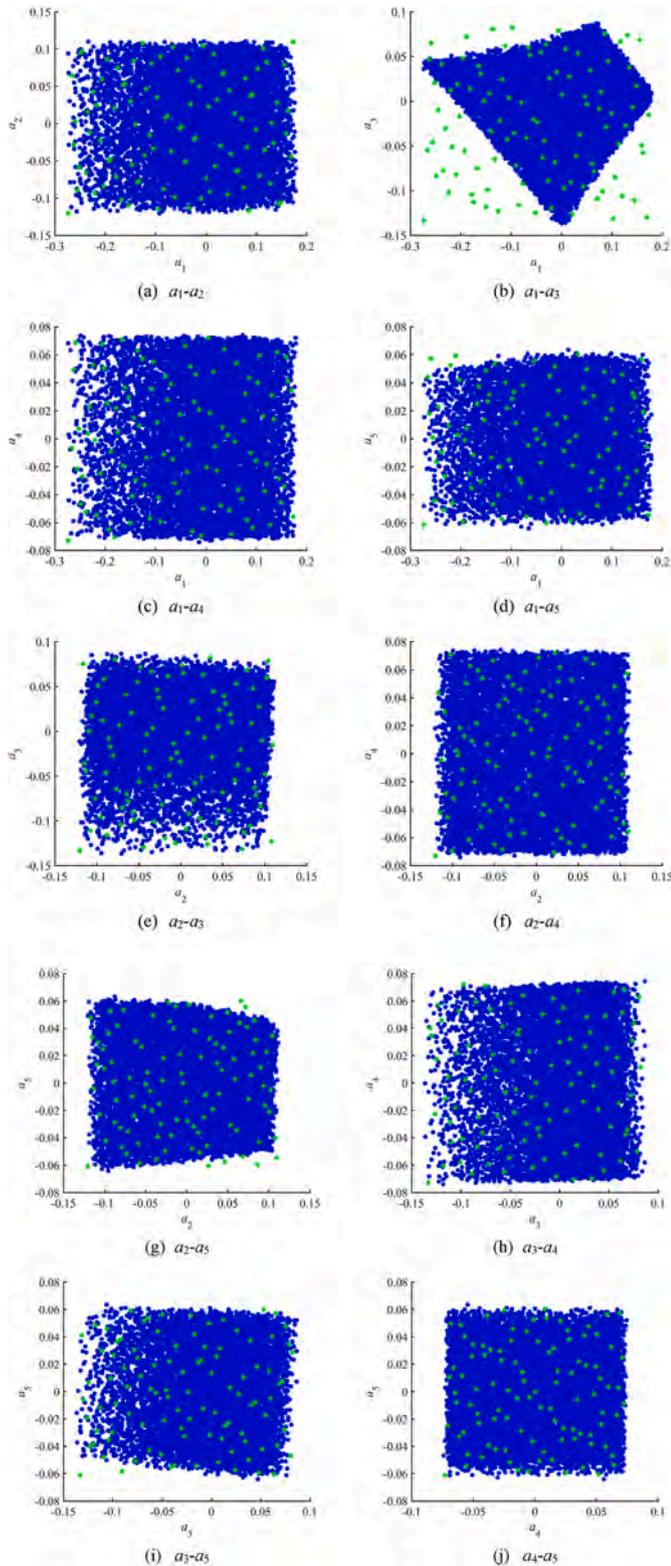


Fig. 8. Relationship among the new design variables.

With no more than 5 basis modes selected, the final dimensionality reduction effect is that the dimensionality of the design space is reduced from 7 to no more than 5. Then, the total drag optimization in the new design space, which is constituted by bases (no more than the first five) that are obtained by dimensionality reduction, can be done.

According to the value range of each variable in the design space after dimensionality reduction obtained above, Optimal Latin

Hypercube Sampling (OLHS) designs (Liu et al., 2018) can be carried out for the new design space with $M = 1, 3, 5,$ and $20, 60, 100$ sample points can be obtained respectively.

It should be noted, however, that the OLHS approach picks samples within a regular hypercube design space, which, based on previous analysis, may be larger than the actual feasible range, but it is difficult to obtain their complex relationship theoretically. From the perspective of practical application, based on the dense sampling with $N = 5000$, 5000 reconstructed new hull forms with their corresponding five new design variable values can be obtained. After doing projections of the five design variables in two-dimensional spaces, we can almost get new design variables' spatial distributions, and a new "boundary" can be regarded as a new constraint among new design variables, although this may lead to a more "conservative" new hull form.

Taking $M = 5$ as an example, the two-dimensional projections of the corresponding basis coefficients of 5000 new samples after dimensionality reduction are respectively given, as shown in the blue dots in Fig. 8. It can be seen clearly that after the dimensionality reduction, the new design space is no longer a hypercube. For the basis modes themselves, however, they do not change depending on whether their coefficient ranges constitute a hypercube. Therefore, if we use OLHS method to obtain the new hulls without considering the constraints, as shown in the 100 green points in Fig. 8, there will be some new hull forms out of the real design space. Predictably, they are beyond the scope of deformation in original design space and may be unrealistic hull forms.

It can be seen from Fig. 8 (b), (g), and (i) that a relatively conservative irregular design space with constraints among design variables can be determined through the "boundary lines" of two-dimensional projections. Constraints can be regarded as linear inequality constraints in the form as follows:

$$\begin{aligned}
 f_1(a_1, a_3) &\leq 0 \\
 f_2(a_1, a_3) &\leq 0 \\
 f_3(a_1, a_3) &\leq 0 \\
 f_4(a_1, a_3) &\leq 0 \\
 f_5(a_2, a_5) &\leq 0 \\
 f_6(a_2, a_5) &\leq 0 \\
 f_7(a_3, a_5) &\leq 0
 \end{aligned} \tag{33}$$

In this new irregular design space, 5-dimensional new design variables with a number of 100 can be obtained through the SMC method with the use of maximum-minimum distance criterion, which fully satisfy all the above inequalities, making full use of the design-space boundaries and avoiding the waste of calculation resources.

Shown above, the constraints between the first five basis coefficients (a_1, a_2, a_3, a_4, a_5) have already been obtained by two-dimensional projections. For the cases of $M = 1$ and $M = 3$, only a_1 and (a_1, a_2, a_3) changes. Specifically, for the case of $M = 1$, only the maximum and minimum values of the projection a_1 need to be given through dense sampling, that is $a_1 \in \left[\min_{x_i \in X} (x_i - \bar{x})^T u_1, \max_{x_i \in X} (x_i - \bar{x})^T u_1 \right], i = 1, 2, \dots, N$. For the case of $M = 3$, after roughly giving maximum and minimum values of (a_1, a_2, a_3) like above, the constraint relationship between the three coefficients needs to be given through two-dimensional projections of $(a_1, a_2), (a_1, a_3), (a_2, a_3)$, which can be seen from Fig. 8 (a), (b) and (e) in the case of $M = 5$, so the projections of $(a_1, a_2), (a_1, a_3), (a_2, a_3)$ are not given repeatedly.

Therefore, after design-space dimensionality reduction, the new hull form optimization problem whose design-space dimensionalities are $M = 1, 3, 5$ respectively, can be defined as follows:

$$\min R_f = R_f + R_w \tag{34}$$

s.t.

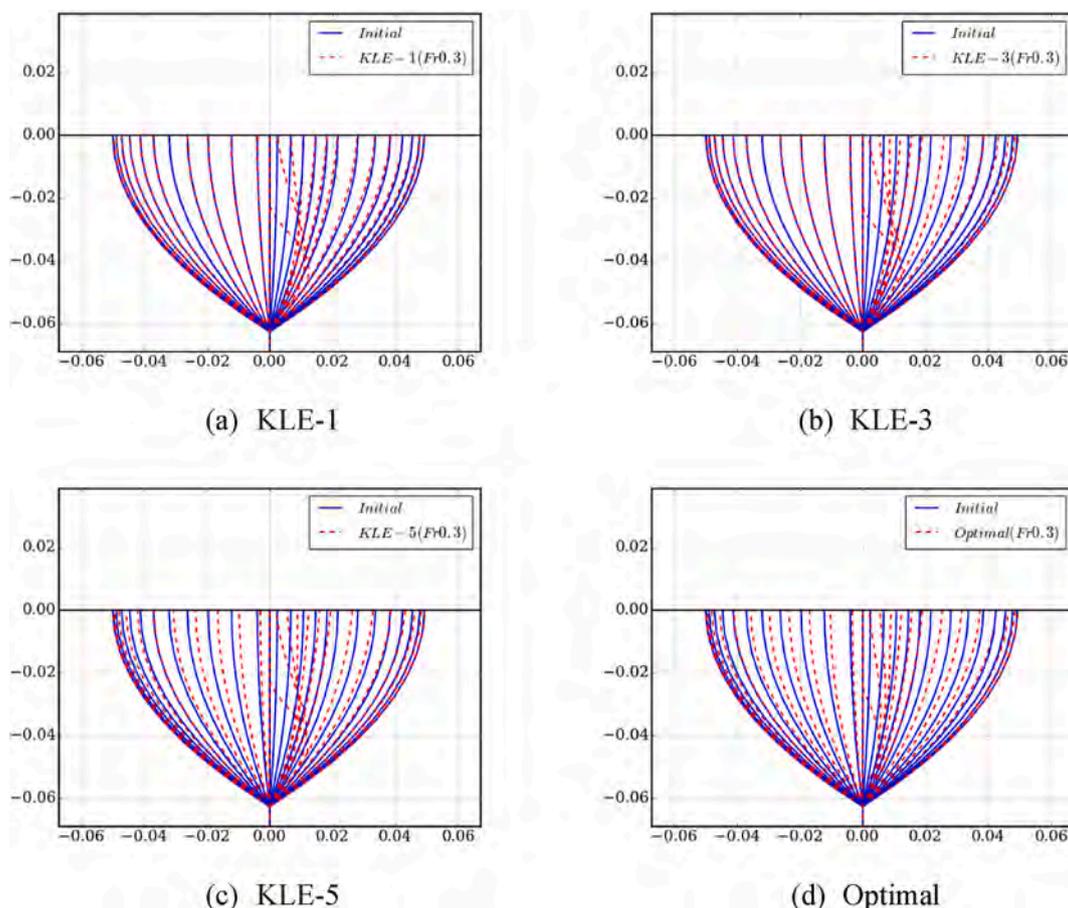


Fig. 11. Hull line comparisons of initial and optimal hulls.

hull in the 5-dimensional design space is closest to the optimal hull in the original design space from the view of the free surface wave elevation.

Fig. 13 shows the comparison of dynamic pressure distribution on the hull surface by NMSHIP-SJTU. It can also be seen that with the increase of design-space dimensionality, the area of high- and low-pressure regions on the hull surface of the corresponding optimal hull decreases gradually. It can be predicted that the optimal hull in the 5-dimensional design space has a fewer total drag than that in the 1-dimensional or 3-dimensional design space.

The viscous CFD solver naoe-FOAM-SJTU is used for further verification of each optimal hull. Table 6 below gives the calm-water total drags of the optimal hulls and the initial hull by naoe-FOAM-SJTU. It can be seen that the optimal hull form with relatively better resistance performance can be obtained in the 5-dimensional design space. For the optimization problems in this paper, the maximum relative changes in the wet surface area and the drainage volume of the hull are 3% and 5%, respectively. It can also be seen from Table 6 that, due to the fact that the optimal hulls all have a generated bulbous bow, the wet surface areas of the optimal hulls are all increased compared with the initial Wigley hull without a bulbous bow; as for the drainage volume, however, since the comprehensive effect by the shifting and RBF method, the KLE-5 and Optimal hulls have a bit smaller drainage volumes. To sum up, for the optimal hulls, their relative changes in wet surface area and drainage volume are all less than 2% and 3%, respectively.

Fig. 14 shows the comparison of free surface wave elevation obtained by using naoe-FOAM-SJTU. It can be seen from Fig. 14 that the decrease trend of the optimal hull's free surface wave elevation is consistent with that of the potential-flow-based calculation results, and with the increase of the design-space dimensionality, the decrease of the corresponding optimal hull's free surface wave elevation is more severe.

Fig. 15 shows the comparison of dynamic pressure distribution on the hull surface obtained by using naoe-FOAM-SJTU. The decrease trend of the area of high- and low-pressure regions calculated by potential-flow and viscous-flow solvers are consistent as a result.

In general, the NMSHIP-SJTU solver can be used to optimize the total drag of the hull form at medium and high speeds, and the optimization results are verified by the viscous CFD solver, showing that the optimization effect of the total drag is obvious. In addition, the design-space dimensionality reduction technique can reduce the huge computational cost of hydrodynamic evaluation and obtain the optimal hull with pretty good resistance performance.

It should be noted that, since the hull form in the design space after dimensionality reduction is basically within the possible deformation range in the original high-dimensional design space, the smoothness of the hull grid mainly depends on the rationality of the deformation method and the selection of design variables in the original design space. The mother ship adopted in this paper is Wigley ship. After generating an initial bulbous bow by RBF method, the modified Wigley ship hull is further optimized. In the process of deformation, the NURBS surface was firstly deformed, which was mentioned in by Liu et al. (2019). Local refinement of the control points of the NURBS surface at the bow part was carried out to ensure that after the bulbous bow was generated, the hull surface mesh converted by the new hull NURBS surface had high quality and good smoothness. Therefore, when we do further optimization of the modified Wigley hull by RBF and shifting deformation methods, the deformed hull surface mesh will still have good smoothness. For actual ship form with complex bow and stern shapes, the amount of grid points on the hull surface may be relatively larger, so the dimensionality of the constructed matrix will increase, which will increase the time required for dimensionality reduction, but the essence of

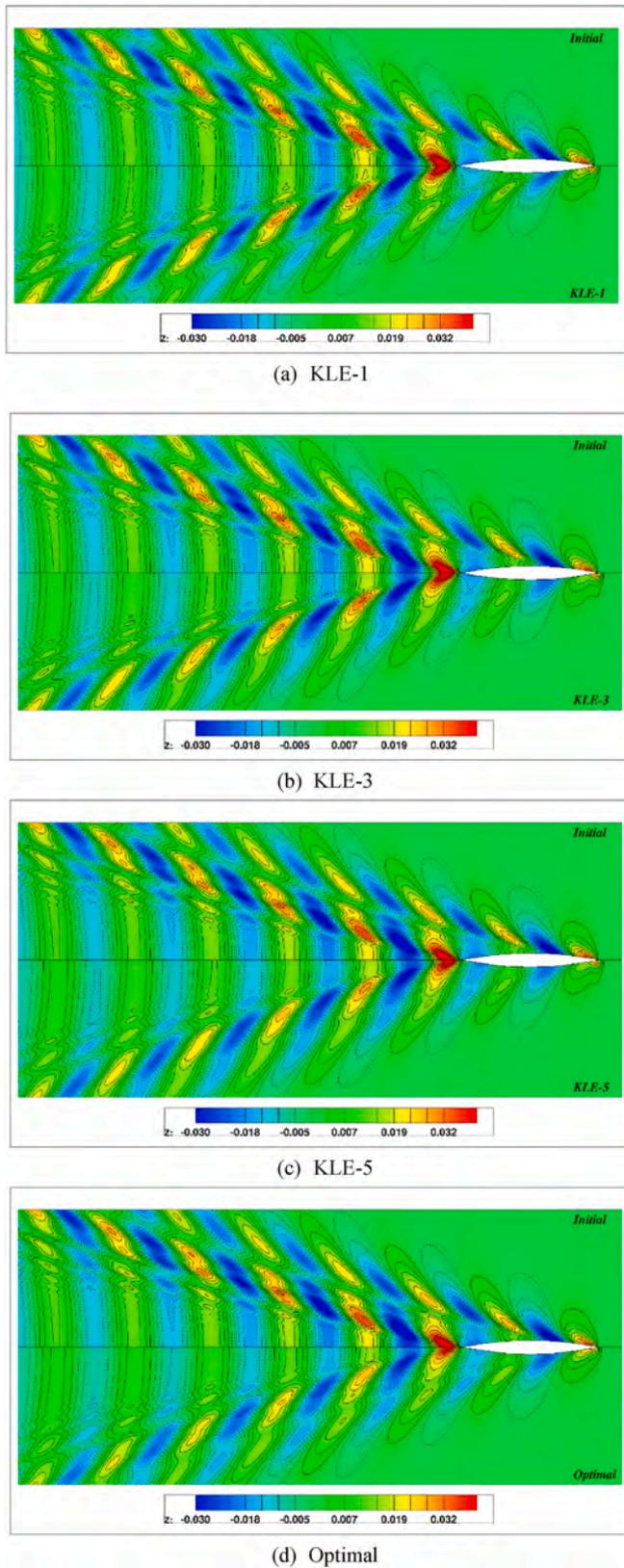


Fig. 12. Wave elevation comparisons of initial and optimal hulls by NMSHIP-SJTU (Unit: m).

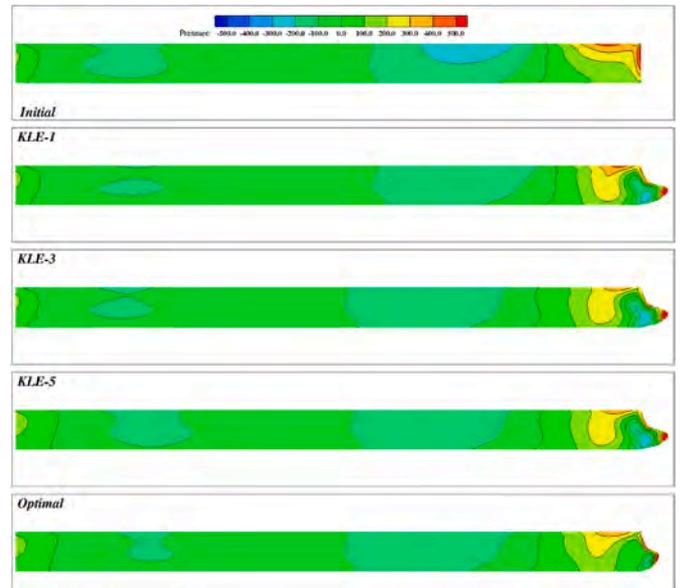


Fig. 13. Pressure comparisons of initial and optimal hulls by NMSHIP-SJTU (Unit: Pa).

Table 6

Comparison of the hydrostatic parameters and total drag of initial and optimal hulls.

Hull form	Static wet surface area S/m^2	Static drainage volume ∇/m^3	Total drag R_T/N	Total drag reduction rate
Initial	2.379	0.177	20.52	–
KLE-1	2.430	0.179	18.75	8.60%
KLE-3	2.430	0.178	18.4	10.30%
KLE-5	2.427	0.176	18.2	11.30%
Optimal	2.397	0.173	17.65	14.00%

the linear reduced order method keeps unchanged.

It can be obviously seen from the body lines of the optimal hulls that the hull grids and NURBS surfaces have good smoothness, which is meaningful to the numerical simulation (hull grids) and the actual construction (hull NURBS surface for IGES file format). What's more, from the view of the corresponding calculated flow fields before and after optimization, since the flow fields are smooth, the calculated hull grids are smooth enough.

Last but not the least, since the 5-dimensional design space can capture 95% deformation information of the 7-dimensional design space, 5% information has been lost. Furthermore, considering that the new design variables (basis coefficients) have some constraints, the 2-dimensional projections are done to see the linear constraints adding into the sampling method and optimization algorithm, however, this operation is relatively conservative. Therefore, the optimal hulls in the 7-dimensional design space and the 5-dimensional design space have some differences in the total drag reduction rate, shown in Table 6. However, in the optimization case with a 7-dimensional design space, 140 sample hulls need to be evaluated to guarantee the fidelity of the Kriging (Krig, 1951) surrogate model, while in optimization cases with 1, 3, 5-dimensional design spaces, only 40, 60, 100 sample hulls are needed. That is to say, the computational cost can have a 28.6% reduction in 5-dimensional design space, which will be much more beneficial in the viscous-flow-based hull form optimization problems, although its optimal hull KLE-5 has a 2.7% total drag difference with the Optimal hull. After comprehensive consideration of optimization effectiveness and efficiency, the new 5-dimensional design space can be regarded as the best design-space dimensionality reduction result.

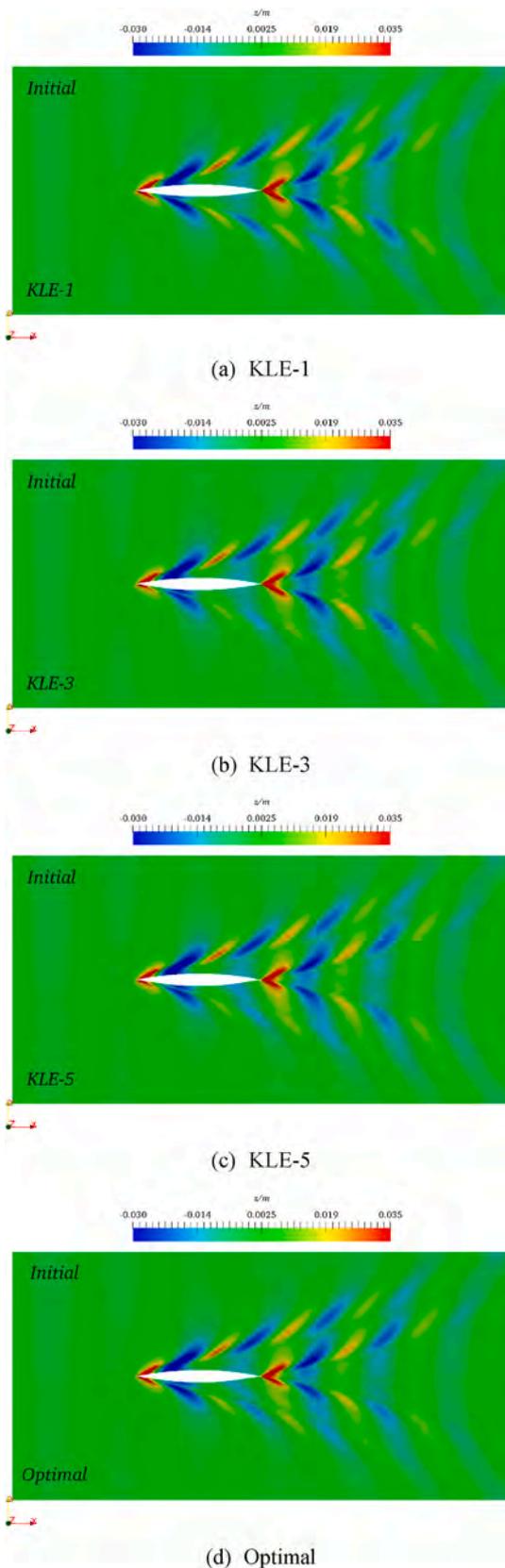


Fig. 14. Wave elevation comparisons of the optimal hulls by naoe-FOAM-SJTU (Unit: m).

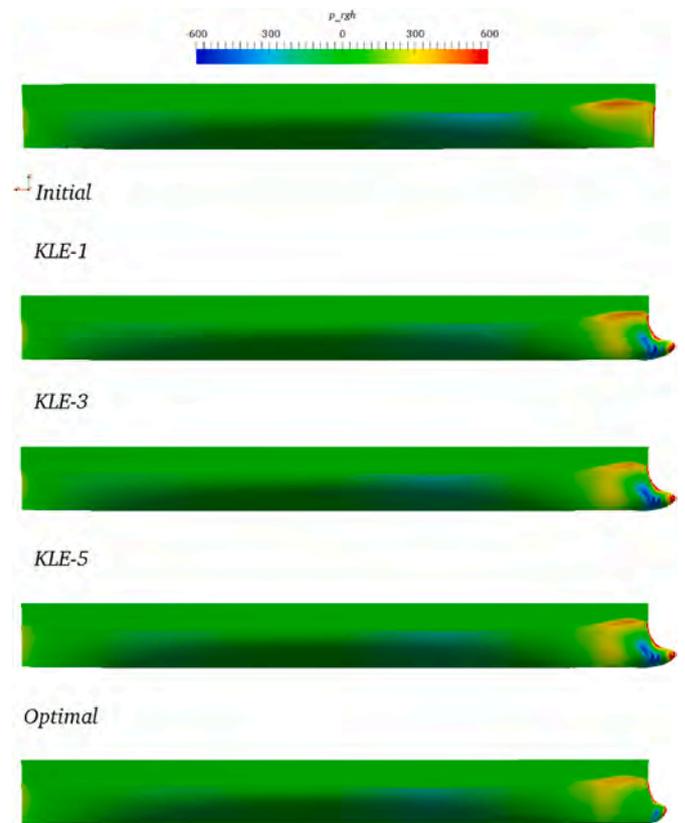


Fig. 15. Pressure comparisons of initial and optimal hulls by naoe-FOAM-SJTU (Unit: Pa).

5. Application of linear dimensionality reduction method in multi-physics field learning in the later stage of hull form optimization

The optimization example in the previous section presents a Simulation-Based Design (SBD) optimization process which is widely used nowadays. There is no doubt that for hull form optimization, the resistance performance is a crucial item for various hydrodynamic performances of ships. However, in SBD optimization process, after doing the hydrodynamic evaluations of each new hull form, only a certain or some performance index values such as total drag, sinkage or trim values are extracted for further optimization to obtain the optimal hull(s), which does not make full use of calculated multi-physics field results of the samples. For instance, for the resistance performance evaluation, the free surface wave elevation and pressure distribution on hull surface can be learned, to analyse the main modes of the flow and the influence of hull form deformation to the flow field, and then to have further understanding of a given hull form deformation design space. Furthermore, according to the existing database of new hull forms, fast hydrodynamic performance prediction, especially the flow field prediction, can also be made for any new hull in the design space.

5.1. Multi-physics field database construction

The basic principle of dimensionality reduction learning for geometric-physical fields is almost the same as the dimensionality reduction techniques used previously in design space. The main difference between the two is that, in the database construction of the geometrical-physical field, the information contained in each new sample hull is not only the variation of grid point coordinates of the hull, but also the wave height of the free surface and the dynamic pressure distribution on the hull surface, so the vector can be written as $x_i^T = (x_1^i,$

$y_1^i, z_1^i, \dots, x_{Np}^i, y_{Np}^i, z_{Np}^i, \eta_1^i, \eta_2^i, \dots, \eta_{Nf}^i, p_1^i, p_2^i, \dots, p_{Np}^i$), where Np, Nf are the grid point numbers of the hull surface and free surface, respectively. When doing database construction, it should be noted that it is best to obtain the dimensionless quantity of the geometric field and each physical field in the scale of the ship model. Specifically, some equations can be used as follows:

$$x = \frac{X}{L_{wt}} \tag{38}$$

$$y = \frac{Y}{L_{wt}} \tag{39}$$

$$z = \frac{Z}{L_{wt}} \tag{40}$$

$$\eta = \frac{H}{U^2/g} \tag{41}$$

$$p = \frac{P}{\rho U^2} \tag{42}$$

There are two main reasons for this. One is to prevent the phenomenon of “large numbers eat the small”: for a ship model sailing in medium or high speed, from a purely numerical point of view, if we all use standard international units to measure, compared with the dynamic pressure on the hull surface, the free surface wave height is quite small. Therefore, if the corresponding vector of each new hull form in the database contains the real values of the multi-physics field, the dimensionality reduction field learning can be completely subject to the pressure distribution, and the information of hull geometrical changes and free surface wave elevation may be ignored, and it’s unlikely to reach the purpose of the entire geometrical-physical field learning. By making all the above physical quantities dimensionless, they can be limited to a very small numerical range, which weakens the possibility of “large numbers eat the small” phenomenon to a certain extent. The other is that it is convenient for physical field data acquisition at different ship model scales: through dimensionless operation to the sample new hulls in the database for geometrical-physical fields, after field learning, the multi-physics field can be quickly forecast for any new hull form within the design space, and the fields are all dimensionless values with respect to the ship model waterline length. If real values of the multi-physics field for a certain ship model length is required, the dimensionless quantities can be converted to real values just according to the above Eqs. 38–42, that is to say, the dimensionless quantity has a wider range of generality, which is not constrained by a certain geometric model scale.

For the 140 new sample hulls of the above modified Wigley ship in the original 7-dimensional design space, the geometric and physical field results obtained by NMSHIP-SJTU are integrated into the multi-physics field learning database. The example grid points contained in the database are shown in Fig. 16, which are required for the potential-flow-based calculation.

Considering the symmetry of the geometrical-physical field about the central shear plane, the calm-water resistance evaluation is carried out by the half-field calculation, as shown in Fig. 16, where the number of free surface grid points is 24460, and the number of hull surface grid points is 8120. Therefore, for the multi-physical field learning of the calm-water resistance optimization database, including hull geometry, hull surface dynamic pressure distribution and free surface wave elevation, the database contains 140 new hull forms, and the number of data points containing multi-physical field information of each new hull is $24460 + 4 \times 8120 = 56940$.

5.2. Influence of mode order on multi-physics field reconstruction effect

Firstly, according to the existing new sample hulls in the database,

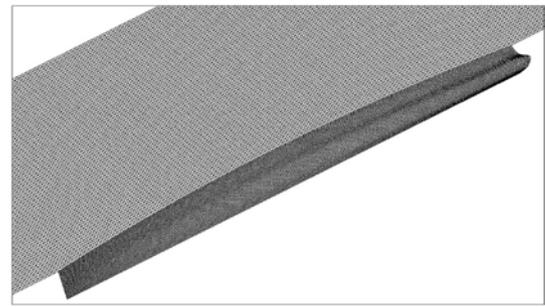


Fig. 16. Calculation mesh example of NMSHIP-SJTU.

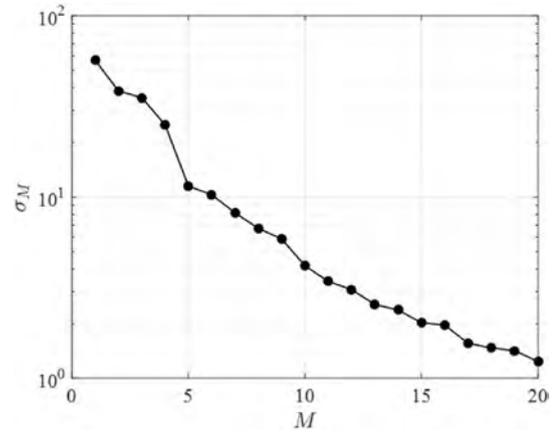


Fig. 17. Changes of eigenvalues corresponding to each mode.

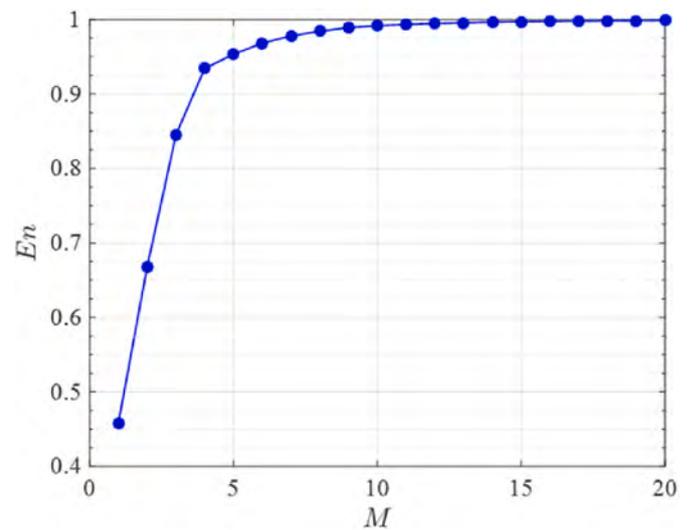


Fig. 18. Changes in the energy proportion En of the first M several modes.

the influence of the mode order M on the flow field reconstruction effect is discussed, and the required mode order can be determined. The changes of arithmetic square root (arranged from largest to smallest) of the eigenvalue and energy proportion corresponding to each mode contained in the database are shown in Figs. 17 and 18 respectively.

As can be seen, in order to capture 99% of the total multi-physics field information, the first 10 modes should be selected. The mean flow field and first 10 basis modes are shown in Fig. 19 below.

It is obvious that with the increase of M , the amplitude of free surface wave height and hull surface pressure corresponding to the basis mode decreases gradually on the whole.

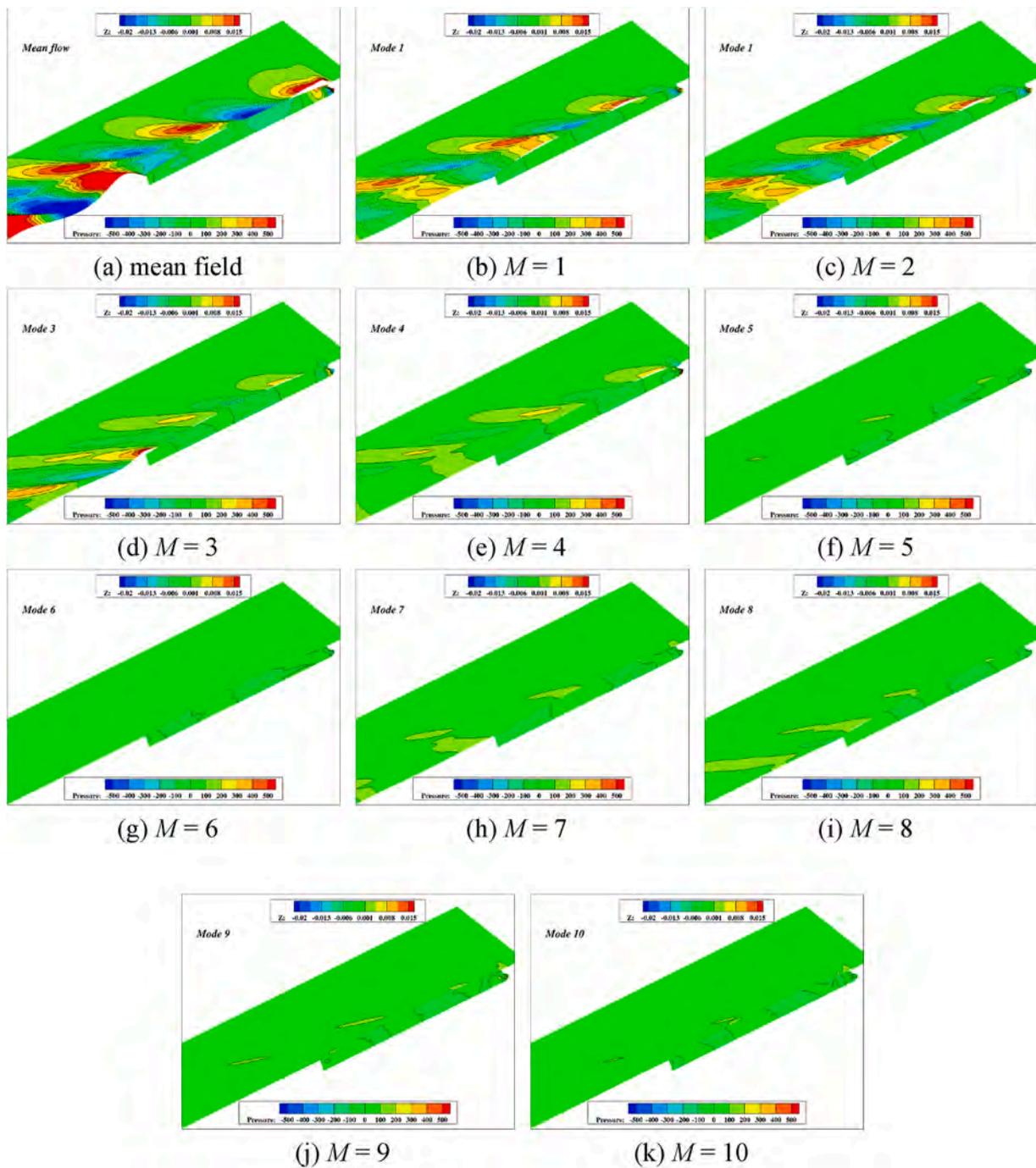


Fig. 19. Mean flow field and first 10 basis modes.

Define relative error e_r for the reconstruction (or prediction) error of the dimensionless multi-physics field $(\tilde{\eta}, \tilde{x}, \tilde{y}, \tilde{z}, \tilde{p})$ after dimensionality reduction and the real field (η, x, y, z, p) of a certain new hull. To be specific, it is the arithmetic square root of ratio of the sum of squares of information difference (total error energy) to the real multi-physics field information (total energy), that is

For each sample in the database, the change relationship between its relative error e_r and M is measured. Six samples are randomly selected from the database, and the relationship is shown in Fig. 20.

It can be found that, with the increase of M , the relative error of the reconstructed field corresponding to each new hull decreases monotonously, and when M reaches 10, the relative error reaches less than

0.5%, indicating that the reconstructed field has high credibility.

Taking Model 3 as an example, the absolute error of the reconstructed flow field with M changing is shown in Fig. 21. It can be intuitively seen that the total absolute error of multi-physics fields decreases gradually with the increase of M .

5.3. Influence of sample number on multi-physics field reconstruction effect

Secondly, according to the existing new sample hulls in the database, the influence of sample number N on the multi-physics field reconstruction effect is discussed to determine whether the number of existing samples is sufficient. Therefore, select 2500 or 5000 sample points in the original design space by Sobol method, which correspond to 2500 or

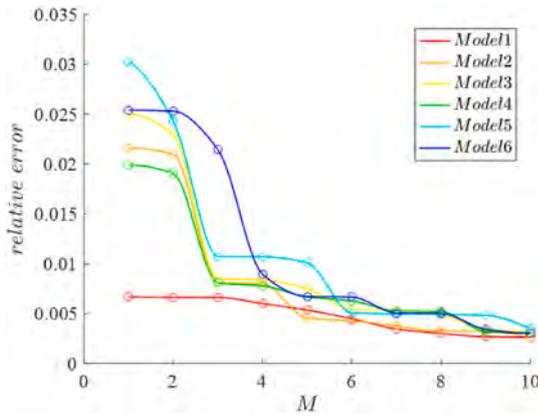


Fig. 20. Relationship between the relative errors of multi-physical fields of each new hull and M .

$$e_r = \sqrt{\frac{\sum_{i=1}^{N_f} (\eta_i - \tilde{\eta}_i)^2 + \sum_{i=1}^{N_p} (x_i - \tilde{x}_i)^2 + \sum_{i=1}^{N_p} (y_i - \tilde{y}_i)^2 + \sum_{i=1}^{N_p} (z_i - \tilde{z}_i)^2 + \sum_{i=1}^{N_p} (p_i - \tilde{p}_i)^2}{\sum_{i=1}^{N_f} \eta_i^2 + \sum_{i=1}^{N_p} x_i^2 + \sum_{i=1}^{N_p} y_i^2 + \sum_{i=1}^{N_p} z_i^2 + \sum_{i=1}^{N_p} p_i^2}} \quad (43)$$

5000 new hull forms, and do the resistance performance evaluations of the new hulls by NMShip-SJTU to build the two new databases, and do the dimensionality reduction multi-physics field learning using the two databases. The variation of E_n with M with different sample numbers N is shown in Fig. 22.

It can be seen that the three change curves almost coincide, indicating that with the increase of M , the change of the proportion of the total energy truncation of multiple physical fields is almost synchronized under different sample numbers. In other words, the database built with 140 new hulls used in the existing hull form optimization case is enough to capture the change rule of the multi-physics fields with the hull form in the original design space.

5.4. Prediction effect analysis of multi-physics field for new hull form

Thirdly, according to the previous analysis, it can be known that for each sample hull existing in the database, if the basis modes and their coefficients corresponding to the sample hull are known, using formula like Eq. (3), the reconstructed multi-physics fields can be quickly obtained. When $M = 10$ is selected, the reconstruction error is small enough. However, more often than not, we need to quickly predict the multi-physical fields of new hulls that are not in the database. Although the basis modes of the database have already been determined, each basis coefficient corresponding to the not-in-the-database new hull cannot be directly solved by formula like Eq. (3). Therefore, it is necessary to find a general method to quickly obtain the basis coefficients.

Considering that each new hull in the database is determined by 7 hull form deformation design variables, and first 10 basis modes with their corresponding coefficients are determined. Therefore, a total of 10 Kriging surrogate models for design variables (hull form deformation parameters) and each basis coefficient (for multi-physics field interpolation) can be established, so that for any new hull, the basis coefficient can be obtained and then the multi-physics field information can be predicted quickly.

The following is a brief verification of the prediction error of the

multi-physics field for a not-in-the-database new hull. Here, variable names ($x_1, x_2, x_3, x_4, x_5, x_6, x_7$) are corresponding to the design variables in Table 2 that define the hull form deformation in the 7-dimensional design space, that is to say, $(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (x, z, y, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

For the convenience of expression, each design variable of the new hull form is unitized, i.e.

$$x_i^* = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}, i = 1, 2, \dots, 7 \quad (44)$$

Obviously, we have $x_i^* \in [0, 1], i = 1, 2, \dots, 7$, since the original design space is a standard hypercube, and define $x^* = (x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*, x_7^*)$.

Take a new hull, whose normalized design variables are $x^* = (0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$, as an example. Through Kriging models, the 10 basis coefficients can be obtained, and the prediction multi-physics field can be got. Compared with the real field evaluated by NMShip-SJTU solver, the physical field information is almost identical shown in Fig. 23, and quantitative analysis of the relative error is given in Table 7.

Therefore, the multi-physical fields corresponding to any new hull in the design space can be quickly predicted with the number of samples N

= 140 and the mode order $M = 10$.

5.5. Sensitivity analysis of hull form optimization design variables

Sensitivity analysis investigates the influence of an independent variable on the dependent variable when other independent variables remain unchanged for a multi-variable problem. By using the above results, the geometrical and flow field corresponding to any new hull can be quickly obtained with high accuracy, and then the influence of all design variables on the flow field details such as free surface wave elevation and hull surface dynamic pressure distribution can be analyzed.

Fig. 24 shows the changing trend of the relative change rate of the wave-making drag coefficient corresponding to the new hull when each dimensionless design variable changes from 0 to 1. The relative change rate of wave-making drag coefficient is positively correlated with the design variables x_3, x_5, x_6, x_7 , but negatively correlated with the design variables x_1, x_2, x_4 . In addition, the design variables x_1, x_3, x_4 have a greater effect on the wave-making drag coefficient, which is called more "sensitive". It can be inferred that the change of design variables x_1, x_3, x_4 will lead to the obvious change of the free surface wave elevation and the dynamic pressure distribution on the hull surface.

The variation trend of free surface wave elevation and the dynamic pressure distribution on the hull surface when changing x_1^* and x_3^* are respectively given below.

Taking 0.1 as an interval, sensitivity analysis is conducted on the design variable x_1^* , and the dimensionality reduction model could be used to quickly obtain the corresponding multi-physics fields of each new hull, as shown in Fig. 25. It can be seen that with the increase of x_1^* , the bulbous bow of the ship elongates, the phase of the wave system moves forward as a whole, and the high- and low-pressure regions of the hull surface move forward with their areas decreasing, which is conducive to the reduction of the wave-making drag coefficient.

Taking 0.1 as an interval, sensitivity analysis is also conducted on the design variable x_3^* . It can be seen from Fig. 26 that with the increase of

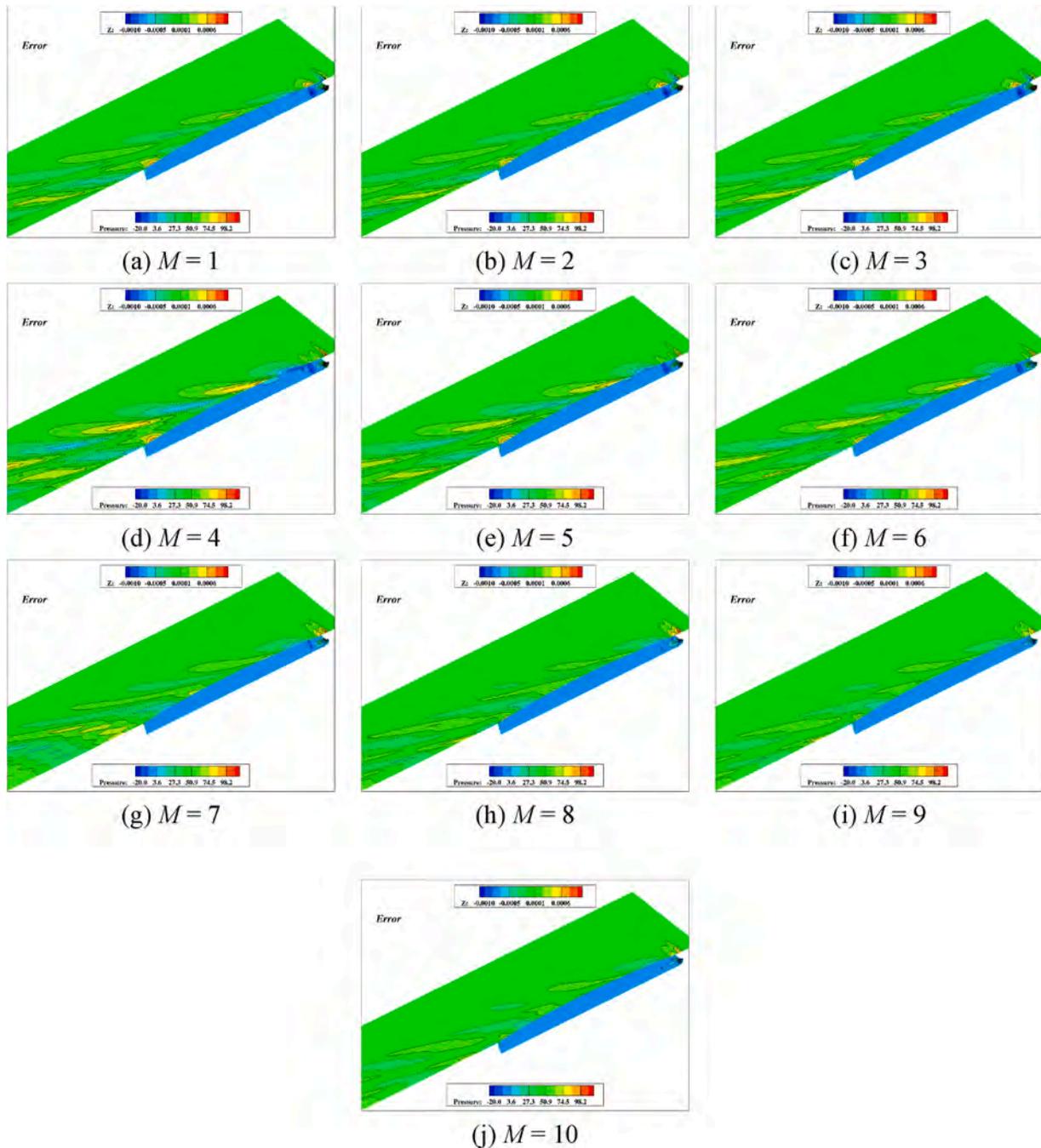


Fig. 21. Absolute error of reconstructed flow field under different M (Unit: m for Z , Pa for Pressure).

x_3^* , the bulbous bow becomes wider, and the phase of the wave systems move forward as a whole. However, the peak value of the bow wave becomes larger, and the area of high-pressure region at the bow increases, which is not conducive to the reduction of the wave-making drag coefficient.

The above conclusions obtained from the flow field are consistent with the sensitivity analysis results in Fig. 24. Similarly, the influence of the other design variables on the flow fields can be analyzed sequentially.

Although the sensitivity analysis in this paper focuses on the wave-making resistance coefficient C_w , which can be directly predicted through Kriging surrogate model, it should be noted that for sensitivity analysis, each new hull form needed is almost located on the boundary surfaces of the design space, so it will hardly be in the database of flow

field learning. That is to say, if there are no flow field learning steps, in order to directly observe the change of hull surface pressure distribution and the free surface wave elevation with respect to a certain design variable, additional numerical simulations should be done, making the computational cost larger. Through flow field learning, the main flow field information corresponding to any new hull form can be obtained within a few seconds, and the overall error can be controlled within 1%. Therefore, it can be considered as high-fidelity and high-efficiency prediction.

It can also be seen intuitively from the Fig. 25 in the paper, for instance, that when x_1 changes alone, how the flow field of the new hull will change, which is predicted by the flow field learning using reduced order method instead of numerical simulation directly. According to the NM potential flow theory, the wave-making resistance coefficient C_w and the flow fields, namely the hull surface pressure distribution and the

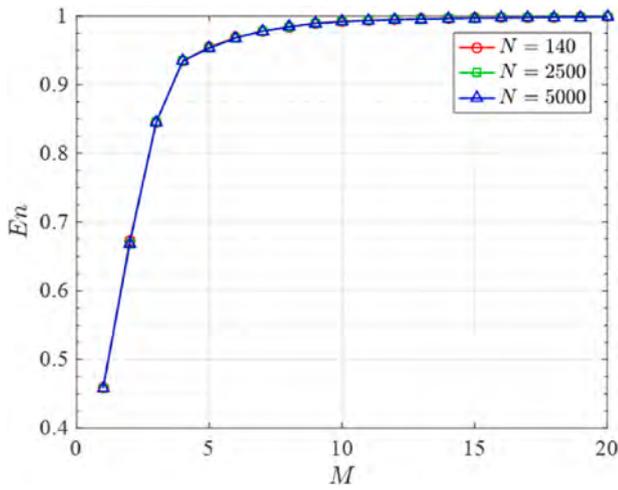


Fig. 22. The variation of En with M with different sample numbers N .

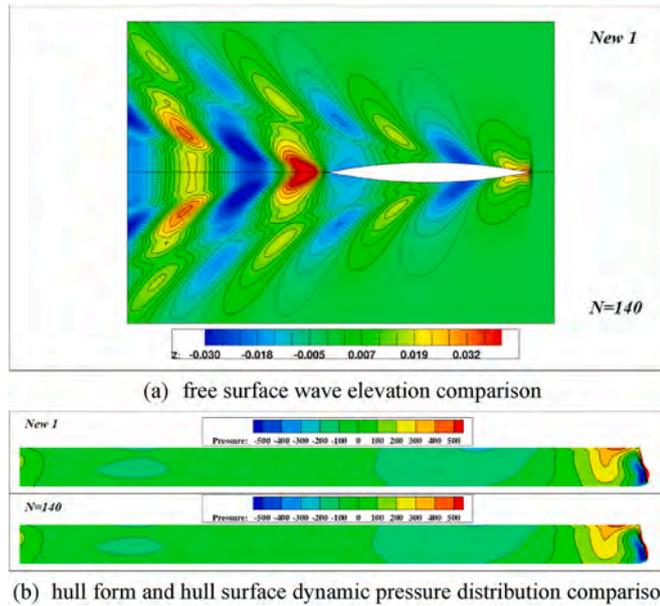


Fig. 23. Comparison between real and predicted flow field by KLE of a new hull.

Table 7
Relative error of each physical field predicted by KLE of a new hull.

Dimensionless field	Real total energy	Total error energy	Relative error e_r
η	12.7929	0.0074	0.0240
x	684.6061	0.0003	0.0007
y	5.3783	0.0002	0.0057
z	11.6820	0.0003	0.0051
p	12.1126	0.0286	0.0486
Sum	726.5719	0.0368	0.0071

free surface wave elevation, are closely related. Therefore, the flow field learning method can be used to match the resistance performance index and the flow field with each other by very small computational costs.

Last but not the least, for other optimization problems, this process can also be implemented in the same way in the later stage of hull form optimization, in order to quickly predict the flow field of any new hull form in the design space of the given optimization problem. This is of greater significance for the high-fidelity hydrodynamic performance

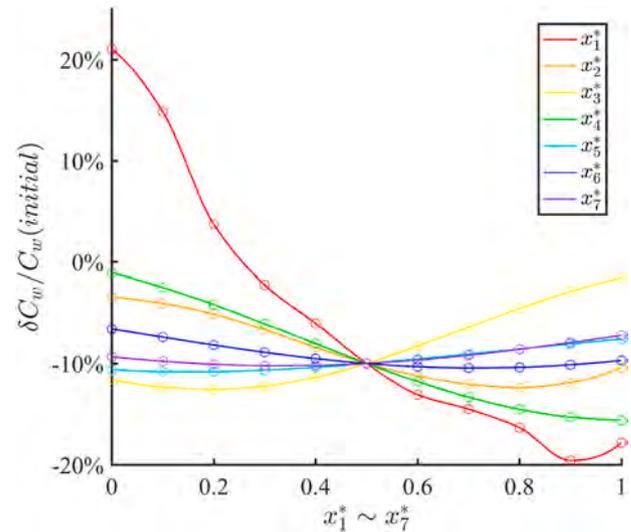


Fig. 24. Sensitivity analysis of design variables to wave-making drag coefficient.

evaluation based on viscous flow, such as using RANS-based solver to obtain the flow fields.

6. Conclusions and future work

In this paper, the POD (or KLE) method is used as the linear dimensionality reduction method, and the steps of dimensionality reduction of the design space are introduced. Furthermore, some important problems for the design-space dimensionality reduction in hull form optimization, such as retainability of fixed control points, irrelevance of the relative order of data to dimensionality reduction results, and decision of the new design space range after dimensionality reduction, are deep discussed. The application of the dimensionality reduction method for design-space dimensionality reduction in the earlier stage of hull form optimization and the multi-physics field learning in the later stage of hull form optimization is given through the resistance optimization of the Wigley ship, and the applicability and reliability of the method are demonstrated by analyzing the influence of mode order and sample number on the reconstruction effect of the hull shape or multi-physics field, and the prediction effect of multi-physics field for new hull form in detail. Current results are promising, showing that the linear dimensionality reduction method can reduce the amount of calculation for surrogate-based optimization, and be used for quick prediction of multi-physics fields of any new form in the design space.

It should be added that the possible deformations in the new design space obtained by the linear dimensionality reduction method cannot completely reach the possible deformations in the original design space due to the information loss of hull form deformation, such as the deformation at the bulbous bow in this example. For any new hull form in original design space, the reconstructed hull form after the dimensionality reduction has more or less error, which is also directly contributing to the slightly inferior drag reduction effect of the optimal hulls in the dimensionality-reduced design space compared with that in the original design space. Future work includes the application extensions of dimensionality reduction method, such as from the linear dimensionality reduction method to the nonlinear dimensionality reduction method, which may make the reconstruction error of new hull form less, and then an optimal hull with better hydrodynamic (resistance) performance may be obtained.

Furthermore, although this paper presents the application of linear dimensionality reduction method in the multi-physics field learning in the later stage of hull form optimization, the objective function of hull

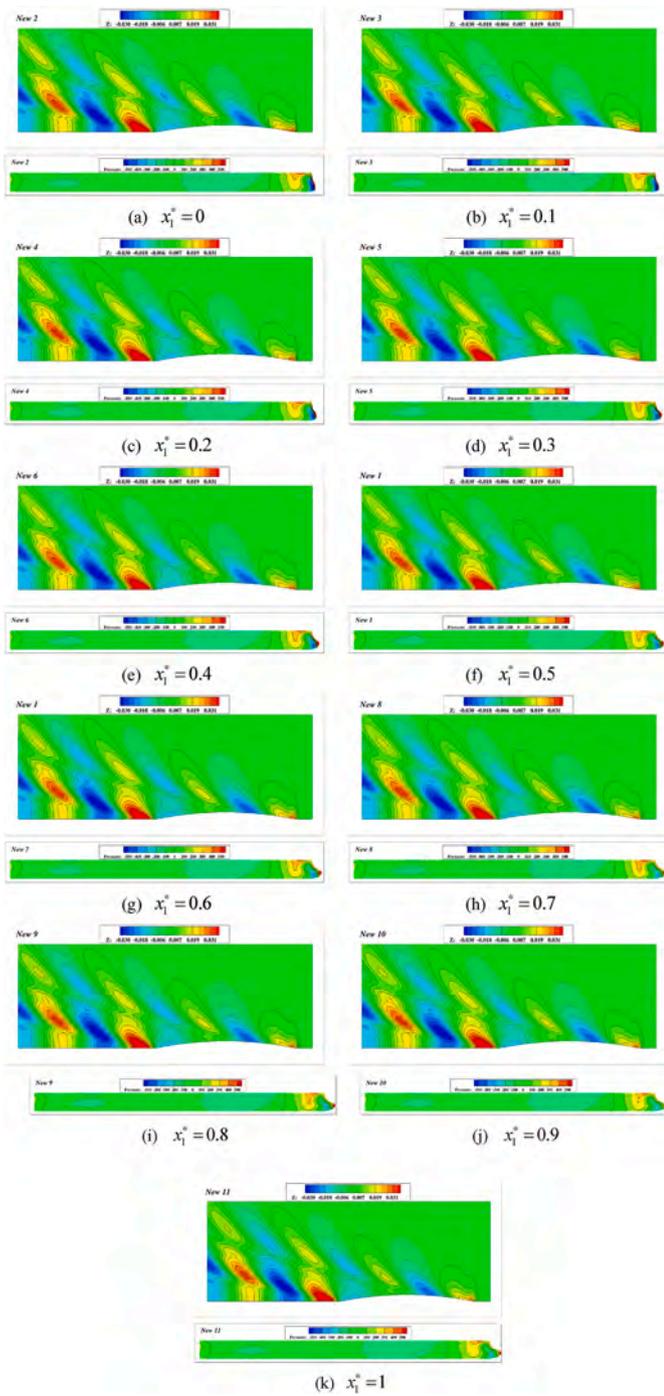


Fig. 25. Variation of flow fields with the first design variable changing.

form optimization cannot be calm-water drag only, and the hydrodynamic performance evaluation method cannot be based on potential flow theory only. In parallel, a similar approach is being applied to optimize the velocity field at the propeller disk by deforming the ship stern with viscous-flow CFD method. Herein, the grid topology relationship at propeller disk of the new hull forms may not be completely consistent, not to say that there is no unified grid topology relationship for unstructured grids.

Financial support

Decheng Wan reports financial support was provided by Shanghai Jiao Tong University.

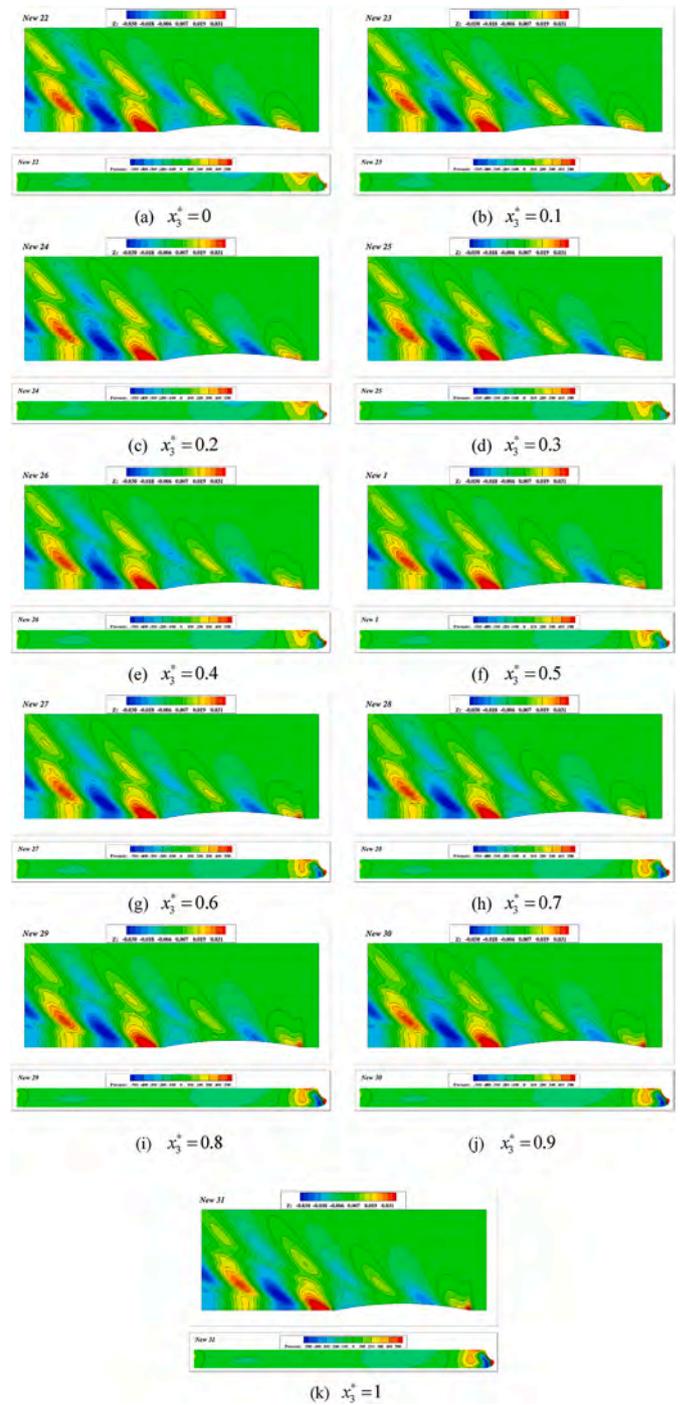


Fig. 26. Variation of flow fields with the third design variable changing.

CRediT authorship contribution statement

Xinwang Liu: Data curation, Writing – original draft, Visualization, Investigation, Software, Validation. **Weiwen Zhao:** Software, Data curation, Visualization, Investigation, Validation. **Decheng Wan:** Supervision, Conceptualization, Methodology, Investigation, Writing – review & editing.

Declaration of competing interest

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

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2019YFB1704200 and 2019YFC0312400), National Natural Science Foundation of China (51879159), to which the authors are most grateful.

References

- Chen, X., Diez, M., Kandasamy, M., Zhang, Z., Campana, E.F., Stern, F., 2014. High-fidelity global optimization of shape design by dimensionality reduction, metamodels and deterministic particle swarm. *Eng. Optim.* 47 (4), 473–494.
- D'Agostino, D., Serani, A., Diez, M., 2020. Design-space assessment and dimensionality reduction: an off-line method for shape reparameterization in simulation-based optimization. *Ocean Eng.* 197, 106852.
- Diez, M., Campana, E.F., Stern, F., 2015. Design-space dimensionality reduction in shape optimization by Karhunen-Loève expansion. *Comput. Methods Appl. Math.* 283, 1525–1544.
- Golchi, S., Campbell, D.A., 2016. Sequentially constrained Monte Carlo. *Comput. Stat. Data Anal.* 97, 98–113.
- Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 504–507.
- Jolliffe, I.T., 1986. Principal component analysis. *J. Appl. Stat.* 87, 41–64.
- Kambhatla, N., Leen, T., 1997. Dimension reduction by local principal component analysis. *Neural Comput.* 9 (7), 1493–1516.
- Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem. Metall. Min. Eng. Soc. South Africa* 52 (6), 119–139.
- Liu, X.W., Wan, D.C., 2020. Hull form optimization of wave-making resistance in different speeds for a luxury cruise ship, 2020 *Chin. J. Ship Res.* 15 (5), 1–10, 40.
- Liu, X.W., Wan, D.C., Chen, G., 2018. Kriging-based surrogate model combined with weighted expected improvement for ship hull form optimization. In: Proceedings of the ASME 37th International Conference on Ocean, Offshore and Arctic Engineering, pp. OMAE2018-78388.
- Liu, X.W., Wan, D.C., Chen, G., Hu, C.H., 2019. Wigley hull form optimization with or without bulbous bow. In: Proceedings of the 29th International Ocean and Polar Engineering Conference, pp. 4486–4493.
- Lloyd, S.P., 1982. Least squares quantization in PCM. *IEEE Trans. Inf. Theor.* 28 (2), 129–137.
- Lukaczyk, T., Palacios, F., Alonso, J.J., Constantine, P.G., 2014. Active subspaces for shape optimization. In: Proceedings of the AIAA Multidisciplinary Design Optimization Conference, p. 1171.
- Luxburg, U.V., 2004. A tutorial on spectral clustering. *Stat. Comput.* 17 (4), 395–416.
- Noblesse, F., Huang, F.X., Yang, C., 2013. The Neumann-Michell theory of ship waves. *J. Eng. Math.* 79 (1), 51–71.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag. A* 2 (11), 559–572.
- Scholkopf, B., 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 10, 1299–1319.
- Serani, A., Diez, M., 2018. Shape optimization under stochastic conditions by design-space augmented dimensionality reduction. In: Proceedings of 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.
- Serani, A., Fasano, G., Liuzzi, G., Lucidi, S., Iemma, U., Campana, E.F., Stern, F., Diez, M., 2016. Ship hydrodynamic optimization by local hybridization of deterministic derivative-free global algorithms. *Appl. Ocean Res.* 59, 115–128.
- Shen, Z.R., Wan, D.C., Carrica, P.M., 2015. Dynamic overset grids in OpenFOAM with application to KCS self-propulsion and maneuvering. *Ocean Eng.* 108, 287–306.
- Sobol, I., 1979. On the systematic search in a hypercube. *SIAM J. Numer. Anal.* 16 (5), 790–793.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometr Intell. Lab. Syst. J.* 1, 37–52.
- Yang, C., Huang, F.X., 2016. An overview of simulation-based hydrodynamic design of ship hull forms. *J. Hydrodyn.* 28 (6), 947–960.