



Multi-fidelity surrogate modeling based on data extension using POD and ANN

Weiji Wang¹, Xuyi Jia¹, Chunlin Gong¹, Chunna Li¹

¹Shaanxi Aerospace Flight Vehicle Design Key Laboratory, School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

Abstract

Surrogate model of flow-field can fast provide flow-field information for various geometries in aircraft conceptual design. Due to the limited sample size of high-fidelity data, it is hard to achieve high prediction accuracy of the flow-field using single-fidelity surrogate models. In this paper, we develop a multi-fidelity modeling method of flow-field, of which the large amount of low-fidelity data can offer trends of the surrogate model, and the small size of high-fidelity data can calibrate model precision. However, flow-field data in various fidelity accumulated in historical designs could be heterogeneous. To handle this problem, we propose a data extension strategy using proper orthogonal decomposition (POD) and artificial neural networks (ANN) to transform the data into the same size. During the modeling, POD extracts low-dimensional features from high-dimensional flow-field data, while ANN maps low-fidelity data to corresponding high-fidelity data. The flow-field prediction of the NACA 4-digit airfoils is carried out to validate the proposed modeling method, and the results indicate that the modeling cost can be obviously reduced in comparison with the modeling purely by small size of high-fidelity data.

Keywords: Multi-fidelity surrogate model; Data extension; POD; ANN; Flow-field prediction

1. Introduction

High-fidelity (HF) flow-field data for various geometries is very important for evaluating aerodynamic performance and aerodynamic shape optimization in the process of aircraft conceptual design. Based on HF flow-field data, we can obtain accurate aerodynamic forces, moments, pressure distributions, and acquire flow-field features as well, such as vortices and shock waves. Traditionally, flow-field data can be obtained by experiments or computational fluid dynamics (CFD). Since experiments are quite expensive and maybe infeasible for hypersonic flight states, CFD simulations are usually used to fetch HF flow-field data. However, CFD simulations require fine meshes and solving complex turbulence models [1-3], which is time consuming for aerodynamic performance evaluation along trajectory, on-line health monitoring and aircraft control.

In recent decades, researchers have begun to explore new alternative methods for the rapid and accurate evaluation of flow fields, particularly focusing on surrogate models that utilize machine learning techniques. In comparison with traditional CFD methods, this method constructs surrogate models by accumulating HF flow-field data, then predicts accurate flow-field data for new geometries or new flight conditions within seconds or even milliseconds. Typical surrogate models for flow-field prediction include deep learning models and reduced-order models (ROMs). Deep learning models construct a direct mapping between input parameters and high-dimensional flow-field through neural networks, such as fully connected neural network (FCNN) [4, 5] and convolutional neural network (CNN) [6, 7]. Due to the high-dimensional feature of flow-field data, the construction of deep-learning models often necessitates the use of complex and extensive neural networks, which presents a significant challenge. In contrast, ROMs tend to find low-dimensional representations of high-dimensional flow-field by dimensionality reduction algorithms, such as proper orthogonal decomposition (POD) [8] and dynamic mode decomposition (DMD) [9]. Then the flow-field can be obtained by simple interpolation methods in the low-dimensional space combined with flow reconstruction in the high-dimensional space. Commonly used interpolation methods include radial basis function interpolation [10], Kriging model [11, 12] and shallow back propagation neural network [13, 14].

Surrogate modeling relies on a large amount of accumulated data from different sources during design in real application. Since the accumulated data varies in fidelity and data in different fidelities are very

likely to be heterogeneous, traditional single-high-fidelity (SHF) surrogate models necessitate numerous computationally expensive HF data to guarantee accuracy. To effectively utilize limited HF data and the accumulated data in different fidelity, researchers have proposed multi-fidelity (MF) surrogate models, in which low-fidelity (LF) data can offer trends of flow-field characteristics and HF data can calibrate model precision [15].

Typical MF surrogate models are categorized into two kinds: one introduces transfer learning to build a fine-tuned network with limited HF data, based on a pre-trained model constructed using sufficient LF data [16, 17]; the other builds surrogate models to map LF data to HF data [18-20]. For the former kind, the effectiveness of transfer learning is affected by the network architecture and the quality of MF data. If the HF and LF data show significant difference, the model parameters of the pre-training model cannot offer meaningful trend for building HF model in the fine-tuning stage. Moreover, most transfer-learning-based MF surrogate models studied are built on homogeneous data in various fidelity, which is not suitable for the heterogeneous data in reality. The latter model can fully utilize flow-field characteristics offered by LF data. But it can only fuse data in two different fidelities, and cannot deal with data in various fidelities and in heterogeneity.

In this paper, we propose a MF surrogate modeling method that is suitable for heterogeneous flow-field data in various fidelities. We use an extension model to extend data into homogeneity, then construct a prediction model using extended homogenous data to predict HF flow-field. The flow-field prediction of airfoil validates the proposed method, and indicates that the method can significantly improve modeling efficiency and reduce total modeling cost by 11.08% with high prediction precision.

2. Multi-fidelity Surrogate Modeling Method

In this section, we first introduce the procedure of the proposed MF surrogate modeling method, then briefly explain the POD and ANN involved in the method.

2.1 The procedure of the developed multi-fidelity surrogate modeling method

The proposed MF surrogate modeling method comprises two main components: the extension model and the prediction model. The extension model transforms LF data into data of the same size as the original HF data, while the prediction model achieves HF flow-field prediction for new geometries. The procedure of the method is presented in Figure 1.

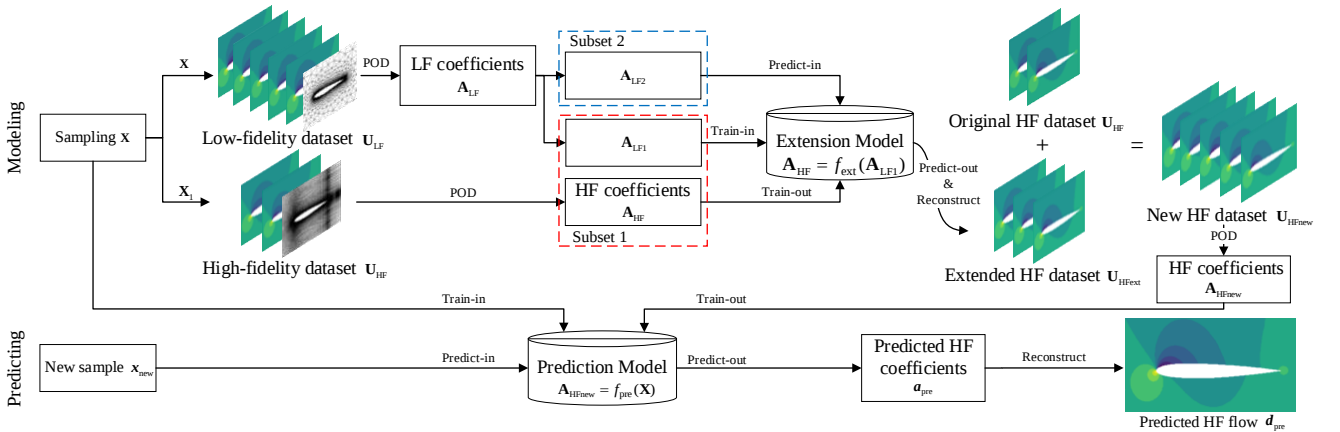


Figure 1 – Procedure of the proposed MF surrogate model.

Details are as follows:

First, the whole sampling plan $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$ is generated by the random Latin hypercube sampling, where $\mathbf{X}$ is for LF samples and $\mathbf{X}_1$ is for HF samples. The ratio of $\mathbf{X}_1$ and $\mathbf{X}$ is defined as MF ratio τ which affect the accuracy and efficiency of MF model. Then the HF dataset $\mathbf{U}_{HF}$ and LF dataset $\mathbf{U}_{LF}$ respectively based on $\mathbf{X}_1$ and $\mathbf{X}$ are computed. In our work, we use the data from the coarse unstructured grid as LF data, and the data from the fine structured grid as HF data. Then two POD models respectively for $\mathbf{U}_{LF}$ and $\mathbf{U}_{HF}$ are built, which are expressed as

$$\mathbf{U}_{LF} = \mathbf{A}_{LF} \cdot \Phi_{LF} + \bar{\mathbf{U}}_{LF} \quad (1)$$

$$\mathbf{U}_{\text{HF}} = \mathbf{A}_{\text{HF}} \cdot \Phi_{\text{HF}} + \bar{\mathbf{U}}_{\text{HF}} \quad (2)$$

Where Φ_{LF} and Φ_{HF} are the mode matrices of the two POD models; $\mathbf{A}_{\text{LF}}$ and $\mathbf{A}_{\text{HF}}$ are the corresponding mode coefficients matrices; $\bar{\mathbf{U}}$ is the mean matrix, where each row of $\bar{\mathbf{U}}$ is the mean vector $\bar{\mathbf{u}}$. The definition of $\bar{\mathbf{u}}$ is given in section 2.2. The mode coefficients for LF data consist of $\mathbf{A}_{\text{LF1}}$ and $\mathbf{A}_{\text{LF2}}$, which are respectively based on $\mathbf{X}_1$ and $\mathbf{X}_2$. These mode coefficients are low-dimensional representations of the flow-field data, thus it is much easier to fuse data in low-dimensional space. Then the entire mode coefficients are divided into two subsets $\mathbf{S}_1$ and $\mathbf{S}_2$, of which $\mathbf{S}_1 = \{\mathbf{A}_{\text{HF}}, \mathbf{A}_{\text{LF1}}\}$ are the mode coefficients from the sampling plan $\mathbf{X}_1$, and $\mathbf{S}_2 = \{\mathbf{A}_{\text{LF2}}\}$ are the mode coefficients from the sampling plan $\mathbf{X}_2$. Since the mode coefficients matrices $\mathbf{A}_{\text{HF}}$ and $\mathbf{A}_{\text{LF1}}$ correspond to each other, we can map $\mathbf{A}_{\text{LF1}}$ to $\mathbf{A}_{\text{HF}}$ by the ANN, which is defined as the extension model $\mathbf{A}_{\text{HF}} = f_{\text{ext}}(\mathbf{A}_{\text{LF1}})$. By using the extension model, we can obtain the predicted mode coefficients matrix $\mathbf{A}_{\text{HFext}}$ of the subset $\mathbf{S}_2$ by

$$\mathbf{A}_{\text{HFext}} = f_{\text{ext}}(\mathbf{A}_{\text{LF2}}) \quad (3)$$

Thus, we can obtain the predicted HF flow-field of $\mathbf{X}_2$ through POD reconstruction $\mathbf{U}_{\text{HFext}} = \mathbf{A}_{\text{HFext}} \cdot \Phi_{\text{HF}} + \bar{\mathbf{U}}_{\text{HF}}$. Subsequently a new HF flow-field dataset $\mathbf{U}_{\text{HFnew}} = \{\mathbf{U}_{\text{HF}}, \mathbf{U}_{\text{HFext}}\}$ is constituted. In general, we can fuse various-fidelity flow-field data in heterogeneity by the same way, and finally obtain a dataset consisting of both original and predicted HF data. Then another POD model based on the new dataset $\mathbf{U}_{\text{HFnew}}$ is built, and the mode coefficients matrix can be computed by solving

$$\mathbf{U}_{\text{HFnew}} = \mathbf{A}_{\text{HFnew}} \cdot \Phi_{\text{HFnew}} + \bar{\mathbf{U}}_{\text{HFnew}} \quad (4)$$

Finally, the prediction model f_{pre} can be constructed by the ANN, which maps $\mathbf{X}$ to $\mathbf{A}_{\text{HFnew}}$, expressed as

$$\mathbf{A}_{\text{HFnew}} = f_{\text{pre}}(\mathbf{X}) \quad (5)$$

In the predicting stage, the mode coefficients vector $\mathbf{a}_{\text{pre}}$ of a new geometry $\mathbf{x}_{\text{new}}$ can be predicted by

$$\mathbf{a}_{\text{pre}} = f_{\text{pre}}(\mathbf{x}_{\text{new}}) \quad (6)$$

Then the HF flow-field $\mathbf{u}_{\text{pre}}$ could be acquired through POD reconstruction

$$\mathbf{u}_{\text{pre}} = \mathbf{a}_{\text{pre}} \cdot \Phi_{\text{HFnew}} + \bar{\mathbf{u}}_{\text{HFnew}} \quad (7)$$

Details of the POD and ANN are briefly introduced in the following sections.

2.2 POD-based flow-field feature extraction

POD is a linear model decomposition method used in dimensionality reduction to reduce the complexity of describing a system, particularly in structures and fluids.

Consider a data matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]^T \in \mathbb{R}^{m \times n}$ with m samples where each sample is an n -dimensional vector $\mathbf{u}_i = [u_1, u_2, \dots, u_n]^T$. The main objective of POD is to find a set of POD modes that represent the directions with maximum variance. The coefficients of these modes are low-dimensional representations of the original high-dimensional data.

To ensure the POD modes capture the directions with maximum variance of the data, zero-centered technique is used

$$\mathbf{u}_k^* = \mathbf{u}_k - \bar{\mathbf{u}} \quad (8)$$

where $\mathbf{u}_k$ is the original k th sample, $\mathbf{u}_k^*$ is the standardized k th sample and $\bar{\mathbf{u}} = \sum_{i=1}^m \mathbf{u}_i / m$. Then the covariance matrix $\mathbf{C} \in \mathbb{R}^{m \times m}$ of the standardized data $\mathbf{U}^* = [\mathbf{u}_1^*, \mathbf{u}_2^*, \dots, \mathbf{u}_m^*]^T \in \mathbb{R}^{m \times n}$ is computed by

$$\mathbf{C} = \frac{1}{m}[\mathbf{U}^* \cdot (\mathbf{U}^*)^T] \quad (9)$$

The covariance matrix summarizes the relationships between different features and their variances. To extract these features, we perform the eigen decomposition and obtain the eigenvectors and eigenvalues of $\mathbf{C}$

$$\mathbf{C} \cdot \mathbf{V} = \mathbf{D} \cdot \mathbf{V} \quad (10)$$

where, $\mathbf{V}=[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]^T \in \mathbb{R}^{m \times m}$ represents the matrix of eigenvectors and $\mathbf{D}=\text{diag}([d_1, d_2, \dots, d_m]) \in \mathbb{R}^{m \times m}$ represents the diagonal matrix composed of all the eigenvalues. Each POD mode ϕ_i corresponds to an eigenvector and the corresponding eigenvalue is

$$\phi_i = \frac{(\mathbf{U}^*)^T \cdot \mathbf{v}_i}{\sqrt{d_i}} \quad (11)$$

Modes with the largest eigenvalues capture the most important features. Sort the modes by eigenvalues in descending order and select the top r modes $\Phi=[\phi_1, \phi_2, \dots, \phi_r]^T \in \mathbb{R}^{r \times n}$ to retain key features. The number of modes r can be chosen using the relative information content (RIC) criterion

$$E_n = \frac{\sum_{i=1}^r d_i}{\sum_{i=1}^m d_i} \quad (12)$$

The RIC indicates the percentage of total variance captured by the selected POD modes. For reduced-order modeling, RIC value of 99.9% or higher are commonly set. The mode coefficient matrix $\mathbf{A}=[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]^T \in \mathbb{R}^{m \times r}$ for the chosen POD modes can be computed by

$$\mathbf{A} = \mathbf{U}^* \cdot \Phi^T \quad (13)$$

Mode coefficient $\mathbf{a}_i$ represents the coordinates of each sample in the POD latent space.

Therefore, the approximate reconstruction of the original i th data $\tilde{\mathbf{u}}_i$ is expressed as

$$\tilde{\mathbf{u}}_i \approx \mathbf{a}_i \cdot \Phi_i + \bar{\mathbf{u}} \quad (14)$$

In this paper, the pressure field is modeled, thus the data matrix is expressed as $\mathbf{U}_p=[\mathbf{u}_{p1}, \mathbf{u}_{p2}, \dots, \mathbf{u}_{pm}]^T$.

2.3 ANN-based flow-field prediction

The Artificial Neural Network (ANN) is a computational model inspired by biological neural networks in the human brain. The backpropagation neural network (BPNN) is one of the most widely used feedforward artificial neural networks, known for its strong nonlinear mapping capability and highly adaptive learning ability, making it suitable for HF flow-field prediction and MF model establishment. A typical BPNN consists of an input layer, one or multiple hidden layers, and an output layer. The k th hidden layer has ns_k neurons. Figure 2 shows the structure of a BPNN with two hidden layers. The neural network maps n -dimensional input vector $\mathbf{x}=[x_1, x_2, \dots, x_n]^T$ to m -dimensional output vector $\mathbf{y}=[y_1, y_2, \dots, y_m]^T$.

The training process of BPNN mainly includes three steps: weight initialization, forward propagation and backward propagation of errors.

Details are as follow:

- (1) Weight Initialization: Initialize the weights matrix ω_{initial} of each neuron using a specific strategy.
- (2) Forward propagation: Process input data $\mathbf{x}$ layer by layer through the network towards the output layer. If there is a difference between the predicted output $\hat{\mathbf{y}}$ and the actual output $\mathbf{y}$, the error is calculated by a loss function, typically mean squared error (MSE)

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (15)$$

- (3) Backward propagation of errors: Propagate the error back through the network, updating the

weights matrix ω_{updated} of each neuron using a learning algorithm to continuously adjust the output closer to the desired value.

For our proposed MF model, the specific meaning of input x and output y depends on the types of models. For the extension model, x represents the mode coefficients of LF data $\mathbf{a}_{\text{LF1}}$, and y represents the mode coefficients of HF data $\mathbf{a}_{\text{HF}}$. For the prediction model, x represents the geometric parameters, and y represents the modal coefficients of the extended HF data $\mathbf{a}_{\text{HFnew}}$.

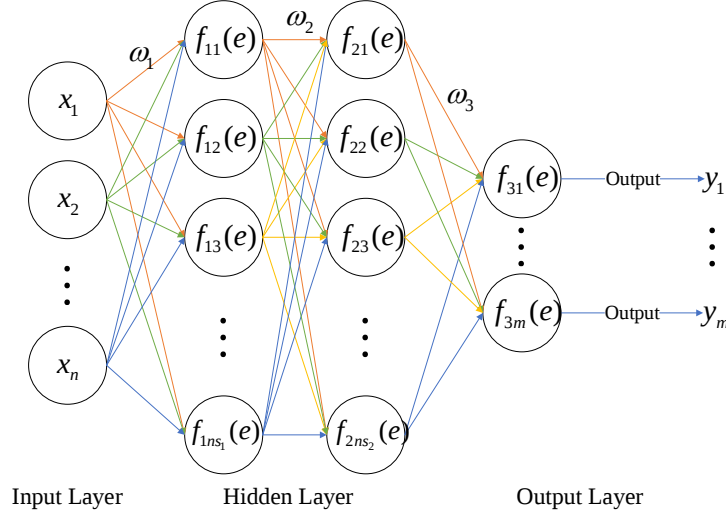


Figure 2 – Structure of BPNN.

3. Subsonic flow-field prediction of the NACA 4-digit airfoils

The developed MF modeling method is validated and assessed by predicting subsonic flow-fields of the NACA 4-digit airfoils. The flow-field data by the Reynold-averaged Navier-Stokes (RANS) solver with fine grid is used as HF data, while the flow-field data by the Euler solver with sparse grid is used as LF data. Then the flow-field model is built under the design condition of $M_\infty=0.55$, $\alpha = 3^\circ$ and $Re=5.8 \times 10^6$. Data acquirement and model construction are both performed on a PC with the Intel Core i7-10750H and NVIDIA RTX 2070.

3.1 Sampling and grids for building the multi-fidelity surrogate model

The NACA 4-digit airfoils are parametrized by the CST method based on a NACA0012 airfoil with six parameters respectively on the upper and lower surface. Thus, each airfoil is described by 12 parameters. The extension of these 12 parameters is specified within $\pm 30\%$ of the parameters of the NACA0012 airfoil. Five hundred samples are generated by the Random Latin hypercube sampling within the extension, of which 400 samples are used for training and the remaining 100 samples are used for testing. The entire design space of the airfoils is illustrated in Figure 3. The whole 400 samples in training set are taken as $\mathbf{X}$ and $\mathbf{X}_1$ is randomly chosen based on the MF ratio τ .

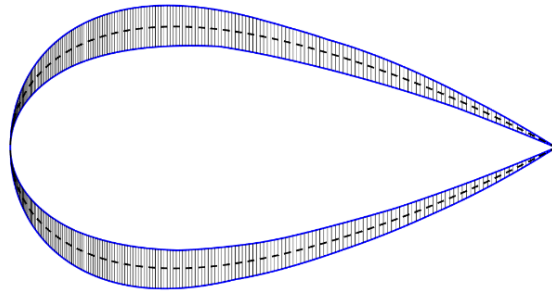


Figure 3 – Airfoil design space (dashed line represents baseline airfoil).

In our work, the heterogeneity of data on different fidelities is reflected on the number of grid cells and the type of grids. Two sets of baseline grids with different levels of sparsity are used, which are respectively shown in Figure 4 and Figure 5. Grids of the remaining airfoils are generated by the RBFs-based mesh deformation technology [21].

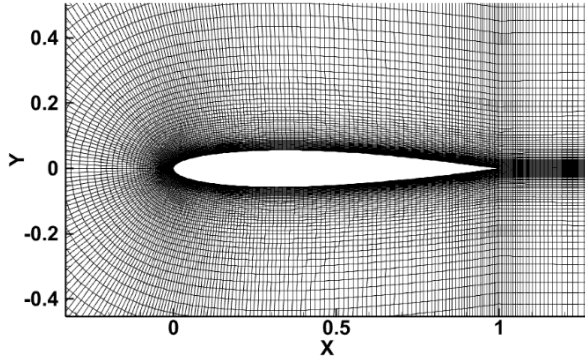


Figure 4–Fine grid (Structured, 82044 cells).

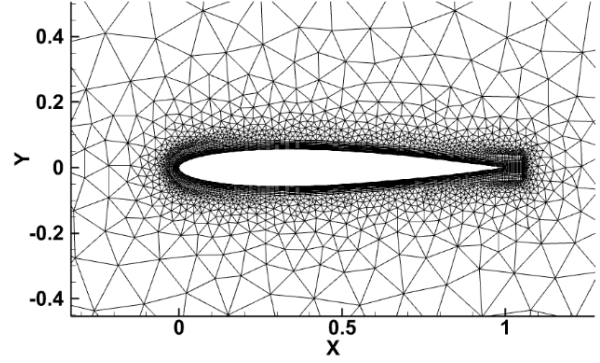


Figure 5–Coarse grid (Hybrid, 7161 cells).

3.2 Setups for CFD simulations

The CFD simulations are performed using the commercial software *Fluent*. Table 1 compares the lift and drag coefficients in different fidelities of the baseline NACA0012 airfoil, and the computational cost required to generate samples for each fidelity level as well.

Table 1 – Comparison of MF data.

Fidelity	Grid size	Grid type	Physics model	cl	cd	Cost/s
High	82,044	Structured	RANS	0.400	0.009	274.26
Low	7,161	Hybrid	Inviscid	0.541	0.008	79.20

The trends of the lift and drag coefficients of the high- and low-fidelity samples in the test set are respectively shown in Figure 6 and Figure 7. Although there is a significant difference between the lift and drag coefficients of HF and LF samples, they generally maintain the same trends of variation. Additionally, because the LF data are obtained using the Euler solver, the friction drag is not considered, thus the LF drag coefficient is much smaller than that from using the RANS solver. Figure 8 compares the HF pressure field with the LF pressure field of the first test sample. It can be seen that the isobars of the HF flow-field are smoother than those of the LF sample due to finer grid.

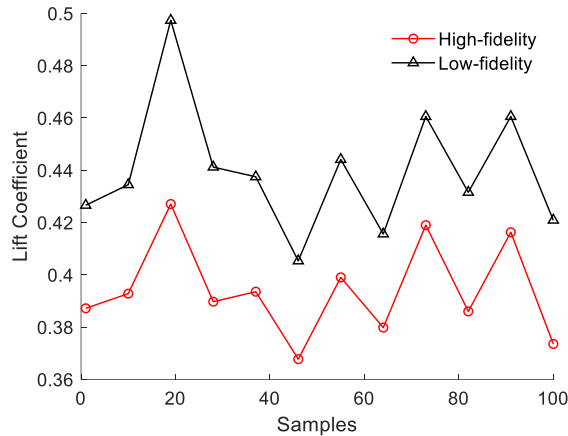


Figure 6 – Lift coefficient comparison.

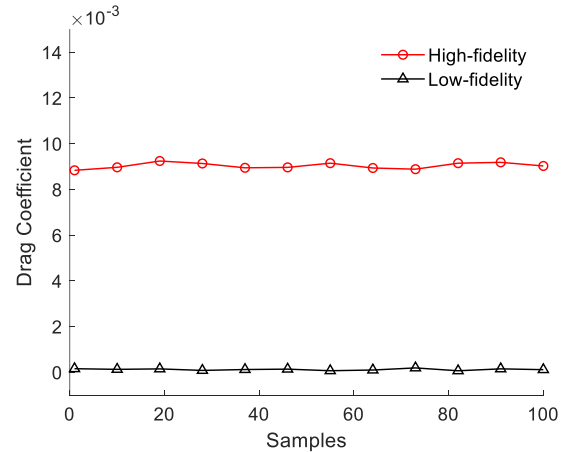
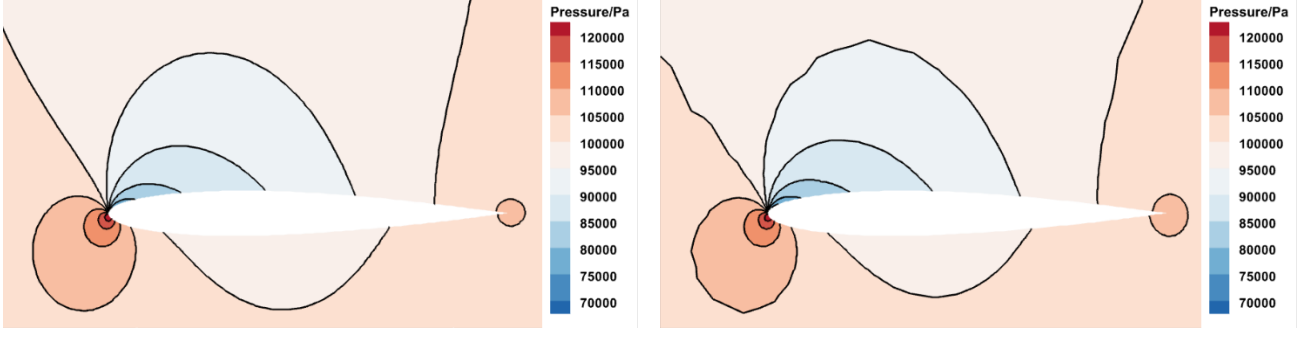


Figure 7 – Drag coefficient comparison.



a) HF pressure field.

b) LF pressure field.

Figure 8 – Comparison between HF pressure field and LF pressure field.

3.3 Error evaluation methods and model settings

The accuracy of the proposed MF surrogate modeling method is qualified by measuring its mean prediction error $\bar{E}$ and maximum prediction error $E_{\max}$ on the test set of samples. The definition of $\bar{E}$ and $E_{\max}$ are given as follows

$$\bar{E}^{(q)} = \frac{\sum_{i=1}^n |e_i^{(k)}|}{n} \quad (16)$$

$$E_{\max}^{(q)} = \sum_{k=1}^n \max(|e_i^{(k)}|) \quad (17)$$

$$\bar{E} = \frac{1}{m} \sum_{q=1}^m \bar{E}^{(q)} \quad (18)$$

$$E_{\max} = \frac{1}{m} \sum_{q=1}^m E_{\max}^{(q)} \quad (19)$$

where superscript q refers to the q th sample in test set; n is the number of vertex in a grid; $e_i^{(k)}$ is the relative error between the predicted pressure and the original pressure at each vertex; m is the number of test samples.

The parameters of the MF surrogate modeling are set as follow. The first fifteen ($r=15$) modes are chosen for dimensionality reduction and data reconstruction. The extension model has two hidden layers, which have equal number of neurons $ns_1 = ns_2 = 16$. While the prediction model has two layers, of which $ns_1 = 10$ and $ns_2 = 16$, respectively. In addition, to evaluate the proposed MF modeling method, we compare its performance with the SHF model. To avoid the influence of the POD models and the neural network architectures on model accuracy, the number of the POD modes are the same for the two models, and the ANN architecture also stays the same. To minimize the impact of random factors in the neural networks, both the MF and SHF models are independently trained ten times.

3.4 Discussions and results

3.4.1 Effect of MF ratio τ

In current analysis, the number of LF samples is fixed at 400, we only change the number of HF samples to change the MF ratio. The mean relative errors and maximum relative errors of the extension model and prediction model on the test set are respectively presented in Figure 9 and Figure 10. It can be seen that the errors both of the prediction model and the extension model tend to stabilize when τ is larger than 0.6, that is, when the number of HF samples exceeds 240.

When τ is smaller, the model error is significantly larger. This is because the error of the extension model is larger when there are fewer HF samples. When τ increases, the accuracy of the extension model improves. The original HF samples can correct some of the errors introduced by the extension

model.

Additionally, the extension model exhibits significantly larger error on the test set in comparison with the prediction model. This is because the HF samples obtained through the extension model have some errors relative to the original HF samples. However, the original HF samples in the sample set plays an important role in error correction during the modeling.

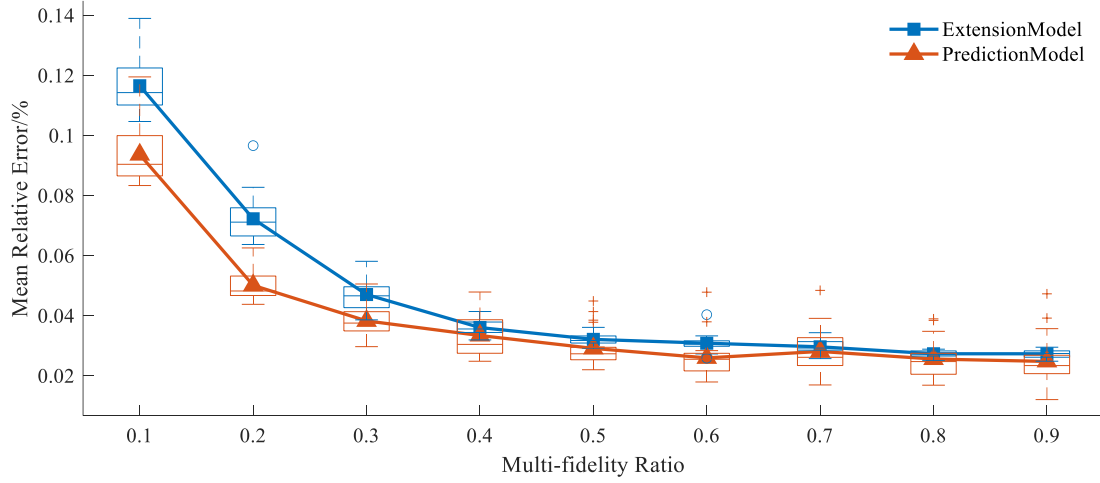


Figure 9 – Mean relative error $\bar{E}$ of the extension model and prediction model on the test set.

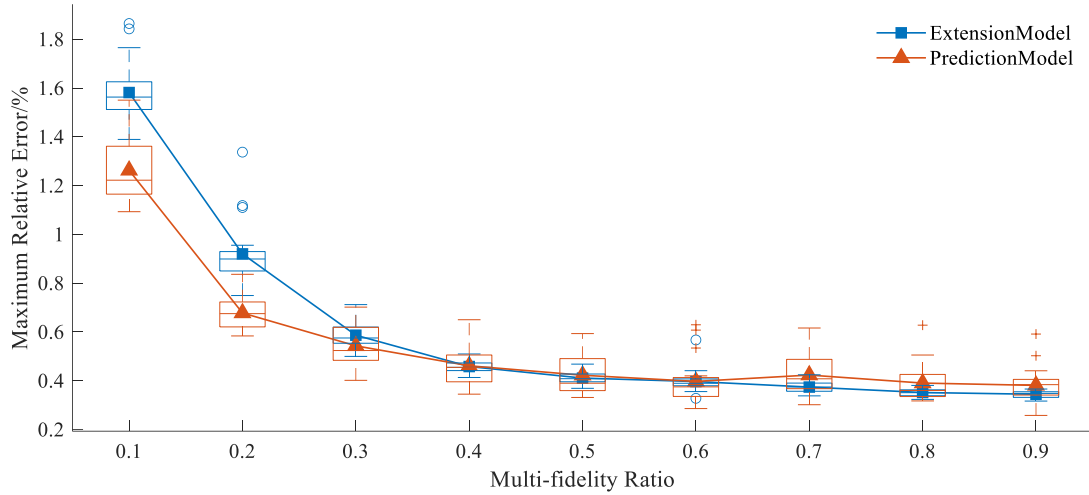


Figure 10 – Maximum relative error $E_{\max}$ of the extension model and prediction model on the test set.

3.4.2 Efficiency comparison with SHF model

To further evaluate the performance of the proposed method, the MF surrogate model is compared with the SHF model constructed purely by the HF samples. The mean relative error and maximum relative error on the test set are respectively compared in Figure 11. It can be seen that, when the number of HF samples is less than 360, with the same number of HF samples, the prediction error of the MF model is smaller than the SHF model. This indicates that introducing LF samples into the MF modelling, can effectively improve model accuracy. When the number of HF samples is 240, the accuracy of the MF model ($E_{\text{Mean}}=0.0260\%$, $E_{\text{Max}}=0.397\%$) is almost the same as that of the SHF model ($E_{\text{Mean}}=0.0253\%$, $E_{\text{Max}}=0.388\%$) established with 400 HF samples. This indicates that under the same prediction accuracy, the MF model can reduce the number of HF samples by about 40%.

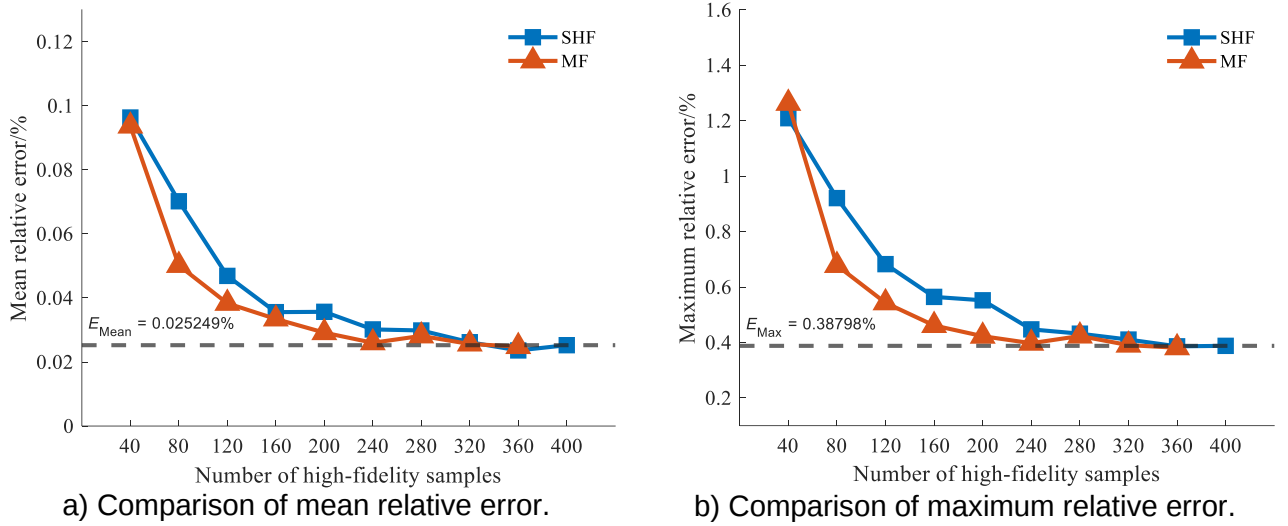


Figure 11 – Curve of relative error for both models as a function of quantity of HF samples.

The total costs of building the MF and SHF models are further compared in Table 2. The cost of the model construction is the average training time for repeatedly training the model ten times. The cost of model construction is the average training time for training ten models repeatedly. From data acquisition to model training, the total cost of the proposed MF model is reduced by 11.08% in comparison with the SHF model. It is worth noting that the time consumptions for solving the high- and low-fidelity subsonic flow-field around airfoils are relatively close. Especially, when the proposed method is applied to predict surface pressure field of 3-D complicated geometries, the MF model will show better performance.

Table 2 – Efficiency comparison of the MF and SHF model.

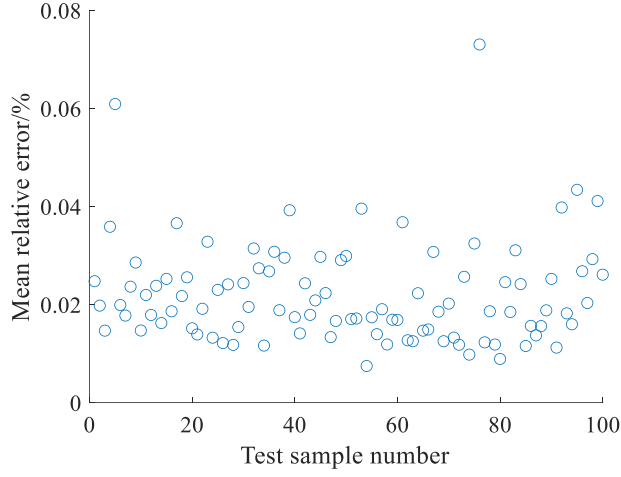
Model	Data acquisition/min	Model construction/s	Total cost/min	Cost reduction/%
SHF	HF: $400 \times 4.57 = 1828.00$	30.70	1828.34	/
MF	HF: $240 \times 4.57 = 1096.8$	Model 1: 24.77	1625.73	11.08
	LF: $400 \times 1.32 = 528.0$	Model 2: 30.80		

*Model 1 refers to the extension model and model 2 refers to the prediction model

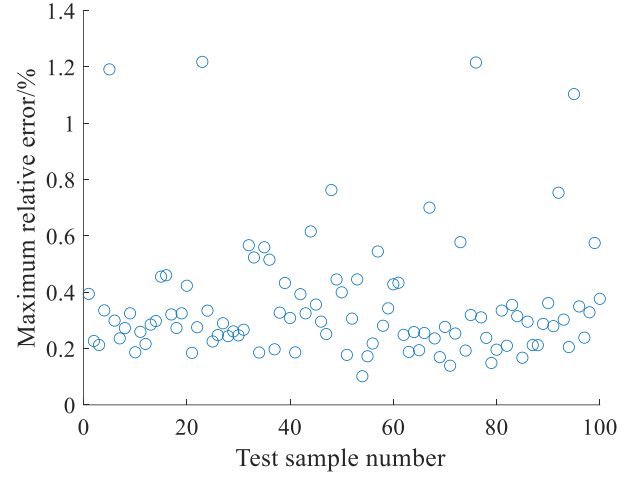
3.4.3 Pressure field prediction by the MF model

To evaluate the accuracy of flow-field prediction by the proposed MF modeling method, the predicted pressure field on test dataset are shown in this section. Based on the analyses in the previous section, a MF ratio of 0.6 is set for establishing the MF model to predict the airfoils' pressure fields. Figure 12 shows the distributions respectively of the mean relative errors and the maximum relative errors for the 100 test samples. Due to the randomness of the neural networks, most of the errors vibrates, and some errors of the test samples are significantly larger. However, the average relative error of all samples remains below 0.08%, and the maximum relative error remains below 1.25%. This indicates that the proposed MF model can predict the flow-field with good accuracy.

The predicted pressure field and absolute error contours of four test samples with the largest relative error are shown in Figure 13. The absolute errors are primarily located near the leading edge and the upper surface. This is because the POD model itself shows errors in the nonlinear regions with large pressure gradients on the upper surface of the airfoil. Therefore, to further improve the accuracy of the MF model, a reduced-order model with higher precision can be adopted.



c) Mean relative error.



d) Maximum relative error.

Figure 12 – Error distribution of proposed MF model on test set.

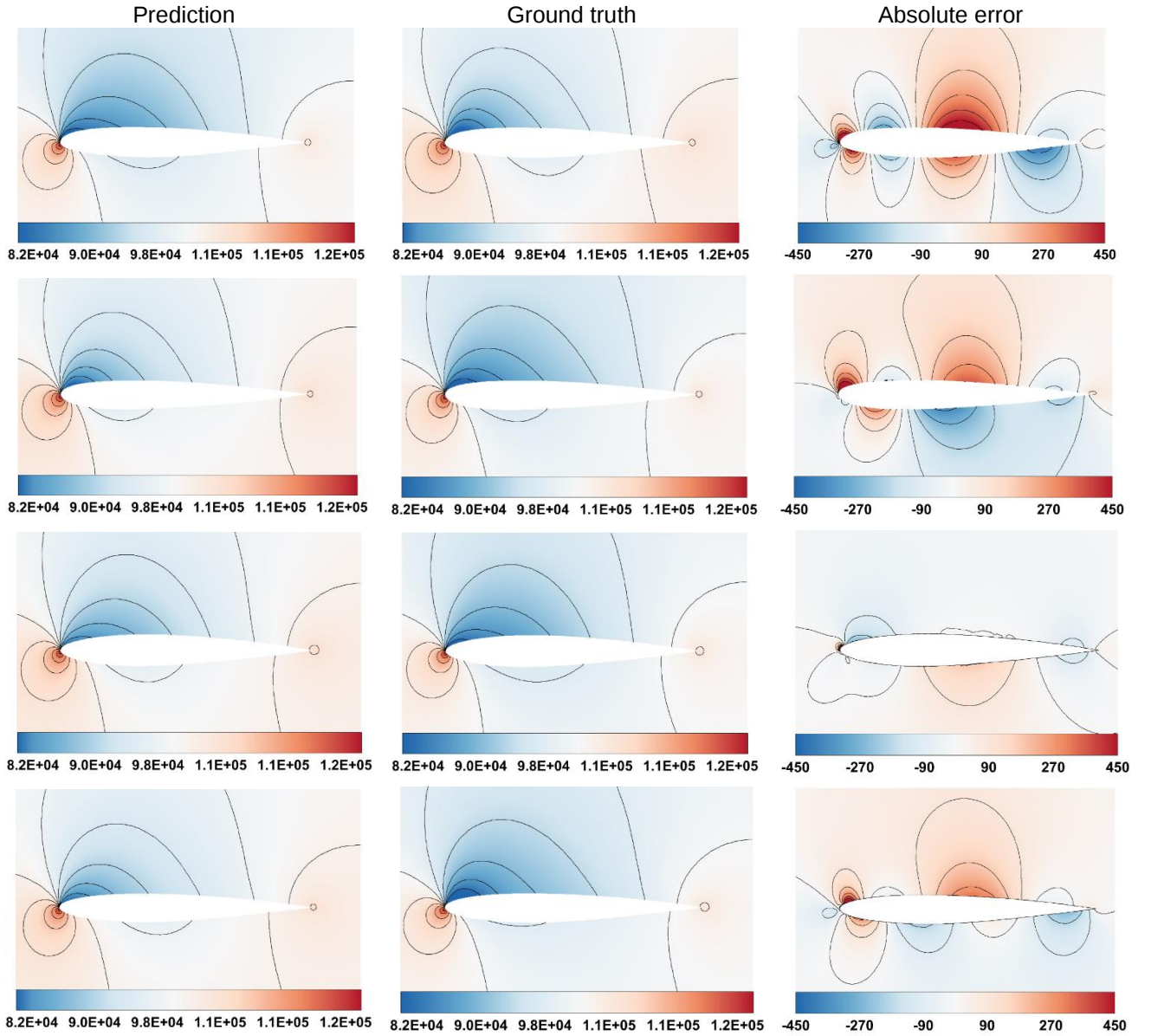


Figure 13 – The predicted pressure field and absolute error contour maps obtained by MF model for the 4 samples with the largest relative error (from top to bottom are sample no.76, no.5, no.23 and

no.95 respectively).

4. Conclusion

In this study, we have developed and validated a MF surrogate model that using POD and ANN to enhance the efficiency and accuracy of aerodynamic data prediction. The construction of the extension model allows for the effective augmentation with limited HF data by enriching it with abundant LF data, thereby reducing the computational cost typically associated with aerodynamic analyses.

Our results demonstrate that the proposed MF model can significantly outperform traditional SHF surrogate models, particularly in scenarios where the available HF data are sparse but there is an abundance of LF data. The proposed MF model can reduce the number of HF samples by about 40% without sacrificing prediction accuracy, thus significantly reducing the reliance on the quantity of HF samples. The analysis on the time consumption shows that the reduction of HF samples saving a total cost of 11.08%. This enhancement in efficiency is valuable for the fast prediction of surface flow-fields on large three-dimensional aircraft. This also demonstrates the practicality of the proposed MF model in real engineering applications.

Future work will focus on refining the integration techniques between the POD and ANN, exploring the application of the model to more complex aerodynamic scenarios, and enhancing the model's adaptability to handle even larger discrepancies in fidelity levels.

5. Acknowledgements

This work is funded by National Natural Science Foundation of China (No. U2141254) and the major advanced research project of Civil Aerospace from State Administration of Science Technology and Industry of China.

6. Contact Author Email Address

Chunna Li: chunnali@nwpu.edu.cn

7. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

- [1] Spalart P. Detached-eddy simulation. *Annual Review of Fluid Mechanics*, Vol 41, pp 181–202, 2009.
- [2] Lesieur M, Métais O. New Trends in Large-Eddy Simulations of Turbulence. *Annual Review of Fluid Mechanics*, Vol. 28, pp 45-82, 1996.
- [3] Speziale C. Analytical Methods for the Development of Reynolds-Stress Closures in Turbulence. *Annual Review of Fluid Mechanics*, Vol. 23, 1991.
- [4] Chang W, Chu X and Fareed A. Heat transfer prediction of supercritical water with artificial neural networks. *Applied Thermal Engineering*. Vol. 131, pp 815-824, 2018.
- [5] Shao C, Liu Y and Zhang Z. Fast Prediction Method of Combustion Chamber Parameters Based on Artificial Neural Network. *Electronics*. Vol. 12, No. 23, pp 47-74, 2023.
- [6] Bublík O, Heidler V and Pecka A. Flow-Field Prediction in Periodic Domains Using a Convolution Neural Network with Hypernetwork Parametrization. *International Journal of Applied Mechanics*. Vol. 15, No. 2, pp 2350018, 2023.
- [7] Peng J, Aubry N and Hua Y. Prediction of internal and external flow with sparse convolution neural network: A computationally effective reduced-order model. *Physics of Fluids*. Vol. 35, No. 2, pp 023605, 2023.
- [8] Berkooz G, Holmes P and Lumley L. The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. *Annual Review of Fluid Mechanics*. Vol. 25, 1993.

- [9] Rowley C, Mezić I, and Bagheri S. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*. Vol. 641, pp 115-127, 2009.
- [10] Qiu Y, Bai J and Hua J. Flow field estimation method based on proper orthogonal decomposition and surrogate model. *Chinese Journal of Aeronautics*. Vol. 34, No. 6, pp 1249-1260 (in Chinese).
- [11] Niu J, Sang W and Li D. Fast Prediction of Multiple Parameters Related to Iced Airfoil Based on POD and Kriging Methods. *Journal of Applied Fluid Mechanics*. Vol. 16, No. 2, pp 325-336, 2023.
- [12] Chen X, Liu L and Long T. A reduced order aerothermodynamic modeling framework for hypersonic vehicles based on surrogate and POD. *Chinese Journal of Aeronautics*. Vol. 28, No. 5, pp 1328-1342, 2015.
- [13] Xie G, Markides C and Sun F. Proper orthogonal decomposition and physical field reconstruction with artificial neural networks (ANN) for supercritical flow problems. *Engineering Analysis with Boundary Elements*. Vol. 140, pp 282-299, 2022.
- [14] San O, Maulik R and Ahmed M. An artificial neural network framework for reduced order modeling of transient flows. *Communications in Nonlinear Science and Numerical Simulation*. Vol. 77, pp 271-287, 2019.
- [15] Wang X, Kou J and Zhang W. Multi-fidelity surrogate reduced-order modeling of steady flow estimation. *International Journal for Numerical Methods in Fluids*. Vol. 92, No. 12, pp 1826-1844, 2019.
- [16] Lyu Y, Zhao, X and Gong Z. Multi-fidelity prediction of fluid flow based on transfer learning using Fourier neural operator. *Physics of Fluids*. Vol. 35, No. 7, pp 77-118, 2023.
- [17] Aliakbari M, Mahmoudi M and Vadasz P. Predicting high-fidelity multiphysics data from low-fidelity fluid flow and transport solvers using physics-informed neural networks. *International Journal of Heat and Fluid Flow*. Vol 96, pp 109002, 2022.
- [18] Conti P, Guo M and Manzoni A. Multi-fidelity reduced-order surrogate modeling. *Proceeding of the Royal Society A*. Vol. 480, No. 2283, pp 20230655.
- [19] Perron C, Rajarm D and Mavris D. Multi-fidelity non-intrusive reduced-order modelling based on manifold alignment. *Proceeding of the Royal Society A*. Vol. 477, No. 2253, pp 20210495, 2021.
- [20] Kou J, Zhang W. Multi-fidelity modeling framework for nonlinear unsteady aerodynamics of airfoils. *Applied Mathematical Modelling*. Vol 76, pp 832-855.
- [21] Boer A, Schoot M and Bijl H. Mesh deformation based on radial basis function interpolation. *Computers & Structures*. Vol. 85, No. 11-14, pp 784-795, 2007.