

Article

A Multi-Fidelity Uncertainty Propagation Model for Multi-Dimensional Correlated Flow Field Responses

Jiangtao Chen, Jiao Zhao, Wei Xiao, Luogeng Lv, Wei Zhao * and Xiaojun Wu

China Aerodynamics Research and Development Center, Mianyang 621000, China; chenjt02@sina.cn (J.C.); zhaoj0828@163.com (J.Z.); xiaov_19@126.com (W.X.); 15280905813@163.com (L.L.); huang7766@sina.com (X.W.)
* Correspondence: ethnuuaa@163.com

Abstract: Given the randomness inherent in fluid dynamics problems and limitations in human cognition, Computational Fluid Dynamics (CFD) modeling and simulation are afflicted with non-negligible uncertainties, casting doubts on the credibility of CFD. Scientifically and rigorously quantifying the uncertainty of CFD is paramount for assessing its credibility and informing engineering decisions. In order to quantify the uncertainty of multidimensional flow field responses stemming from uncertain model parameters, this paper proposes a method based on Gappy Proper Orthogonal Decomposition (POD) for supplementing high-fidelity flow field data within a framework that leverages POD and surrogate models. This approach enables the generation of corresponding high-fidelity flow fields from low-fidelity ones, significantly reducing the cost of high-fidelity flow field computation in uncertainty propagation modeling. Through an analysis of the impact of uncertainty in the coefficients of the Spalart–Allmaras (SA) turbulence model on the distribution of wall friction coefficients for the NACA0012 airfoil and pressure coefficients for the M6 wing, the proposed multi-fidelity modeling approach is demonstrated to offer significant advancements in both accuracy and efficiency compared to single-fidelity methods, providing a robust and efficient prediction model for large-scale random sampling.

Keywords: uncertainty quantification; multi-fidelity model; multidimensional correlated responses; machine learning; flow field reduction



Citation: Chen, J.; Zhao, J.; Xiao, W.; Lv, L.; Zhao, W.; Wu, X. A Multi-Fidelity Uncertainty Propagation Model for Multi-Dimensional Correlated Flow Field Responses. *Aerospace* **2024**, *11*, 263. <https://doi.org/10.3390/aerospace11040263>

Academic Editor: Kung-Ming Chung

Received: 27 January 2024

Revised: 16 March 2024

Accepted: 18 March 2024

Published: 28 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous advancements in mathematical models, numerical algorithms, mesh technologies, and high-performance computing, CFD has emerged as a pivotal tool in numerous critical engineering fields, including aerospace, energy and power, transportation, and beyond. Nevertheless, the credibility of CFD has been a subject of contention due to the non-negligible uncertainties inherent in its models, parameters, and numerical solutions [1]. A comprehensive and rigorous quantification of these uncertainties is imperative for assessing and enhancing the credibility of CFD.

Parameters serve as a significant source of uncertainty in CFD. Due to limitations in human cognition or the inherent randomness in fluid problems, model parameters or inflow conditions may possess a certain degree of uncertainty. To quantify the uncertainty introduced by parameters, scholars have developed various methods, such as Monte Carlo-type random sampling and polynomial chaos [2]. While Monte Carlo-type methods are straightforward, they require a large number of samples to obtain stable and accurate statistical results. Polynomial chaos methods rely on orthogonal polynomial expansions, with coefficients obtained through numerical integration or regression analysis. However, as the dimension of input parameters and the order of expansion increase, the number of samples required for polynomial chaos methods grows drastically, leading to the “curse of dimensionality”. For complex engineering problems, each CFD simulation is computationally expensive, limiting the practical application of these methods due to high computational

costs. In recent years, scholars have attempted to introduce machine learning algorithms into the field of numerical simulation uncertainty quantification, using surrogate models to replace complex simulation systems and combining experimental design methods to reduce the demand for sample data [3–6]. This represents a successful amalgamation of intelligent learning algorithms and uncertainty quantification (UQ) research.

The accuracy of using surrogate models for UQ analysis heavily relies on the accuracy of the surrogate model itself, while the efficiency is determined by the time taken to train and run the surrogate model. Factors such as the size and quality of the training data, feature selection, and the design and training of machine learning algorithms can all affect the accuracy and efficiency of surrogate models. To further improve these aspects, scholars have proposed multi-fidelity modeling methods that leverage the respective strengths of high/low-fidelity analysis models. These methods use a larger number of low-fidelity samples to reduce computational complexity, while also incorporating a smaller number of high-fidelity samples to ensure the predictive performance of the multi-fidelity approximate model. This effectively balances the trade-off between the predictive performance of the approximate model and modeling costs. Commonly used multi-fidelity models include co-Kriging models [7–9] and multi-fidelity neural networks [10,11], which have demonstrated great potential in complex equipment optimization design and provide new ideas for quantifying parameter uncertainty propagation in CFD.

However, most existing multi-fidelity models are only applicable to single-output scenarios. In CFD, the outputs of interest not only include individual variables but may also involve flow field variables that vary with time or space, such as wall pressure coefficient distributions and the unsteady aerodynamic forces of aircraft. There may exist potential correlations among flow field variables at different locations or times, and the dimensionality of the output can reach hundreds or even thousands. Moreover, the response dimensions under different fidelities may differ, such as flow fields under different grid scales or unsteady aerodynamic forces under different time steps. Developing multi-fidelity models suitable for uncertainty propagation modeling remains a challenging problem that urgently needs to be solved.

To address the uncertainty quantification problem in multidimensional and correlated flow fields, we developed a modeling method for uncertainty propagation based on Proper Orthogonal Decomposition (POD) and surrogate models. By employing POD, several key basis functions representing flow structures are identified, reducing the multidimensional flow field responses, which span hundreds or even thousands of dimensions, to a reduced-dimensional representation of approximately ten dimensions, which is composed of basis function coefficients. Subsequently, a prediction model is established between the uncertain model parameters and the basis function coefficients. Since POD provides a bidirectional representation between flow field variables and basis function coefficients, it enables the prediction of flow fields under arbitrary model parameters, providing a reliable prediction model for Monte Carlo and other random sampling methods. However, previous studies have primarily relied on data from a single fidelity level. To fully leverage CFD calculation data from different fidelities, this paper presents a high-fidelity data completion method based on Gappy POD, further reducing the sample requirements of existing methods for high-fidelity data.

The paper is organized as follows. The second section introduces the implementation framework and main theoretical methods of the entire approach, including POD, Gappy POD, and Kriging models. The third section presents the cases studied: the low-speed flow around an NACA0012 airfoil and the transonic flow around an M6 wing. The fourth section contains the analysis results of the cases, including the prediction accuracy analysis of Gappy POD and the entire method, the statistical information of the flow field, and the parameter sensitivity results. Finally, the conclusion and future directions for further research are discussed.

2. Uncertainty Propagation Model Framework and Key Algorithms

In the field of CFD, using different grid densities, time step sizes, etc., can generate computational data of varying fidelities. To reduce the overall modeling cost of the uncertainty propagation model, this study adopts a multi-fidelity modeling framework that combines flow field reduction and surrogate modeling. The overall implementation process is shown in Figure 1. The left part of the figure represents the main algorithms used in each stage, and the right part is the model training process. Below, we focus on introducing the model training process.

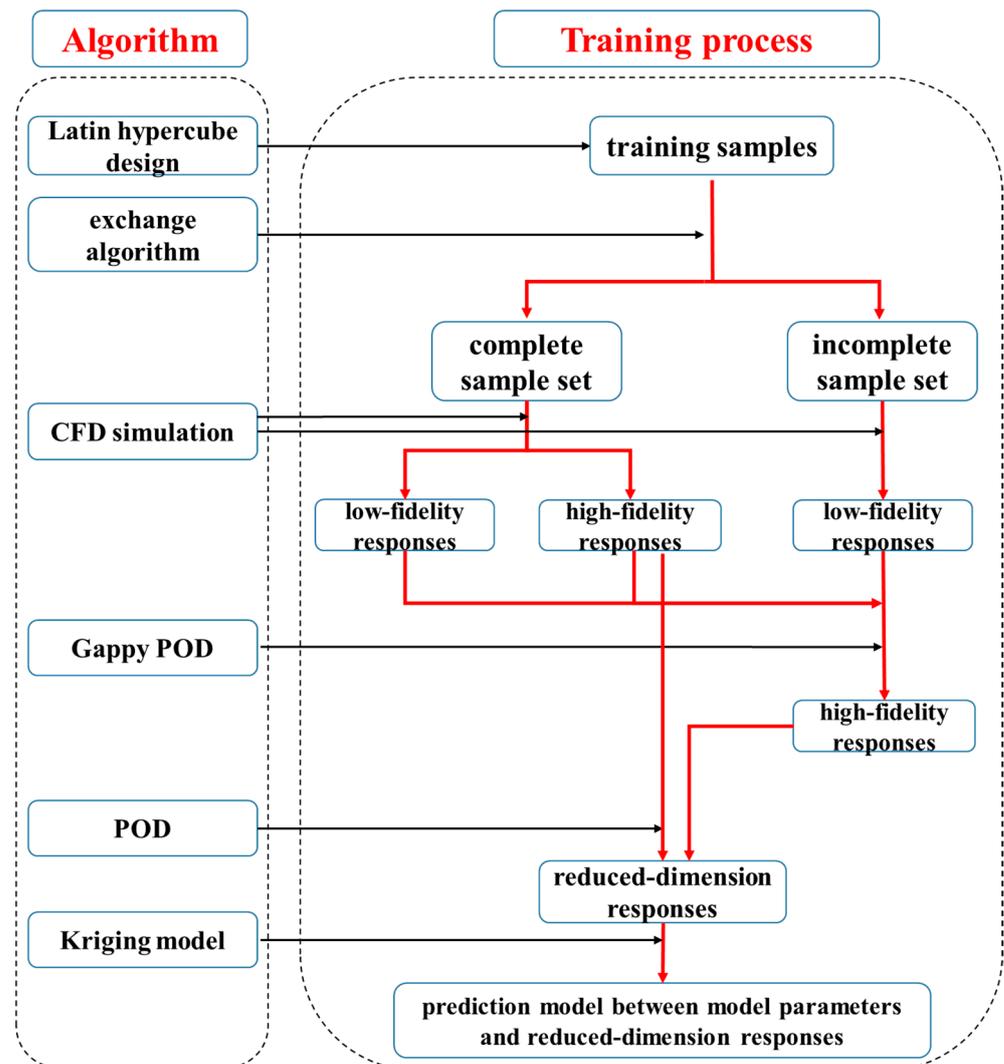


Figure 1. Overall flow chart for multi-fidelity uncertainty propagation model.

2.1. Model Training Process

The entire model training process can be implemented using the following six steps:

- (1) In the input parameter space, Latin hypercube sampling is employed to generate the inputs for the n_{train} training samples.
- (2) Based on the inputs of the training samples, the exchange algorithm [12] based on the Morris–Mitchell criterion [13] is used to divide the n_{train} training samples into two parts: M complete samples and $n_{train} - M$ incomplete samples.
- (3) For the M complete samples, their corresponding high- and low-fidelity outputs are obtained through CFD calculations. For the $n_{train} - M$ incomplete samples, only their corresponding low-fidelity outputs are obtained through CFD calculations, while their high-fidelity outputs are considered unknown. Of course, we also obtain the high-

fidelity outputs of n_{train} incomplete samples through CFD calculations. However, these data are not used for model training but only for testing the prediction ability of the Gappy POD method described below.

- (4) Using the Gappy POD method [14,15], the high-fidelity outputs of n_{train} incomplete samples are predicted. At this point, the high-fidelity output results for all n_{train} training samples can be obtained. The low-fidelity output results of these samples will no longer be used hereinafter.
- (5) POD [16,17] is performed on the high-fidelity outputs of the n_{train} training samples to obtain an orthogonal basis function space. Through projection, the basis function coefficients for each training sample are obtained, which are considered the new outputs.
- (6) Based on the n_{train} training samples, a Kriging model [18,19] is constructed between the input parameters and the basis function coefficients. Since the orthogonal basis functions obtained through POD are mutually orthogonal, an individual model can be constructed for each basis function coefficient. When given new input parameters, the corresponding basis function coefficients can be predicted based on the models, and the complete flow field response can be reconstructed using the bidirectional expression of POD.

2.2. Model Testing Process

In the input parameter space, Latin hypercube sampling is again employed to obtain the inputs for the n_{test} training samples. The CFD program is run to obtain high-fidelity sample outputs, which are used to evaluate the prediction accuracy of the overall model.

2.3. Exchange Algorithm

After obtaining n_{train} training samples, we select M samples from them to perform complete high- and low-fidelity simulations using CFD. Theoretically, the M samples should ideally have good space-filling property in the parameter space. Here, the criterion proposed by Morris and Mitchell [13] is adopted to measure the space-filling property of the sample set, i.e., a good space-filling design should maximize the minimum inter-site distance:

$$\max : \min_{1 \leq i < j \leq n} d_{ij}$$

where d_{ij} is the Euclidean distance between the two samples x_i and x_j , and n represents the number of samples in the sample set.

To achieve a better space-filling property, the exchange algorithm proposed by Cook and Nachtsheim [12] is employed. Specifically, M samples are randomly selected from the training sample set as the complete sample set X_e , with the remaining samples forming the incomplete sample set X_r . The minimum inter-site distance of X_e is calculated. We then exchange the first sample in X_e (denoted as $x_e^{(1)}$) with each of the samples in X_r and retain the exchange that maximizes the minimum inter-site distance of X_e . The process is repeated for each sample in X_e ($x_e^{(2)}, \dots, x_e^{(M)}$).

It should be emphasized that the exchange algorithm involves a degree of randomness, necessitating multiple repeated trials to verify its reliability and stability.

2.4. POD

POD is a powerful mathematical technique used for analyzing and reducing the complexity of high-dimensional systems. This method is widely applied in various fields such as fluid dynamics, structural dynamics, and signal processing, enabling the efficient extraction of important characteristics, the reduction in computational costs, and the enhancement of simulation efficiency. POD provides a valuable tool for the analysis and optimization of complex systems.

The POD method was proposed by Lumley [16], and its basic principle is to identify a new set of bases for the space spanned by a group of vectors in high-dimensional space,

such that the projection of the original vectors onto the set of basis functions is as large as possible. It involves projecting data onto a set of orthogonal basis functions, allowing for the extraction of dominant modes and features.

In this paper, the input and output of the k -th sample are denoted as θ_k and s_k separately. The output of M samples, $\{s_k\}_{k=1}^M$, constitutes a snapshot set. The average value and fluctuation part of the sample are defined as:

$$\begin{aligned}\bar{s} &= \frac{1}{M} \sum_{k=1}^M s_k \\ s'_k &= s_k - \bar{s}\end{aligned}\quad (1)$$

The POD algorithm finds a set of optimal orthogonal bases $\{\boldsymbol{\varphi}_i | i = 1, 2, \dots, N\}$, such that the projection error of the fluctuation part of sample set $\{s'_k\}_{k=1}^M$ in the space spanned by $\Phi = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_n]$ is minimized; that is to say, $Q = \frac{1}{M} \sum_{k=1}^M (\|s'_k - \Phi \boldsymbol{\alpha}^k\|)$ is minimized. Here, $\boldsymbol{\alpha}^k$ is the vector composed of orthogonal polynomial expansion coefficients.

To improve the efficiency and accuracy of POD, Sirovich [17] proposed the snapshot method for the efficient extraction of orthogonal basis vectors. This method assumes that the basis functions are linear combinations of sample snapshots, i.e.:

$$\boldsymbol{\varphi}_i = \sum_{k=1}^M a_k^i s'_k$$

Constructing the snapshot covariance matrix C with element C_{ij} is the inner product of s'_i and s'_j , that is:

$$C_{ij} = \frac{1}{M} (s'_i, s'_j), 1 \leq i, j \leq M$$

Solving its eigenvalues and eigenvectors, M non-negative eigenvalues, $\lambda_i (i = 1, 2, \dots, M)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$, and corresponding eigenvectors, $\mathbf{b}^i (i = 1, 2, \dots, M)$, are obtained. The optimal orthogonal base $\boldsymbol{\varphi}_i$ can be expressed by $\boldsymbol{\varphi}_i = \sum_{k=1}^M b_k^i s'_k$.

In the context of POD analysis, the significance of a basis mode is gauged by the magnitude of its eigenvalue. The generalized energy of the first l basis modes is formulated as:

$$E_l = \left(\sum_{k=1}^l \lambda_k \right) / \left(\sum_{k=1}^M \lambda_k \right) \quad (2)$$

The generalized energy criterion is commonly used to retain the prevailing basis modes, specifically considering the first l basis modes whose generalized energy just exceeds a predefined threshold as the dominant basis modes. Upon determining the dominant basis modes and truncating the basis function space, any given sample snapshot can be approximated as a linear combination of these prevailing basis modes, expressed as:

$$s_k \approx \bar{s} + \sum_{i=1}^l \alpha_i^k \boldsymbol{\varphi}_i \quad (3)$$

where $\boldsymbol{\alpha}^k = (\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k)^T$ is obtained through least-squares method, which means:

$$\boldsymbol{\alpha}^k = (\Phi^T \Phi)^{-1} \Phi^T (s_k - \bar{s})$$

Through POD, the original high-dimensional response is reduced to an l -dimensional response, effectively alleviating the challenge of constructing a surrogate model. In subsequent surrogate modeling, the output is no longer the original response s_k , but a low-

dimensional vector $\vec{\alpha}$ composed of basis mode coefficients. It is noteworthy that Formula (3) facilitates a bidirectional process, enabling not only the dimensionality reduction in the known s_k to obtain $\vec{\alpha}$ but also the utilization of $\vec{\alpha}$ to reconstruct the original output response s_k .

2.5. Gappy POD

The Gappy POD method enables the generation of corresponding high-fidelity flow fields from low-fidelity ones, avoiding time-consuming high-fidelity CFD calculations.

Regarding the i -th sample in the complete sample set, we denote its model input parameters as θ_i , its low-fidelity output vector as $s_i^L = (s_{1,i}^L, s_{2,i}^L, \dots, s_{n_L,i}^L)^T$ with a dimension of n_L , and its high-fidelity output vector as $s_i^H = (s_{1,i}^H, s_{2,i}^H, \dots, s_{n_H,i}^H)^T$ with a dimension of n_H . The high- and low-fidelity outputs are combined into a multi-fidelity sample snapshot denoted as $s_i = \begin{pmatrix} s_i^L \\ s_i^H \end{pmatrix}$ with a dimension of $n_L + n_H$.

Classical POD decomposition is performed for the multi-fidelity snapshots $\{s_i\}_{i=1}^M$. Based on the generalized energy criterion, truncation is performed to obtain an orthogonal space $\Phi = [\varphi_1, \varphi_2, \dots, \varphi_n]$, which is composed of n orthogonal basis functions. Therefore, any multi-fidelity snapshot can be represented as:

$$s \approx \bar{s} + \sum_{j=1}^n \alpha_j \varphi_j$$

where $\bar{s} = \frac{1}{M} \sum_{i=1}^M s_i$. The basis mode coefficient vector $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ is obtained using least-square methods.

For the j -th sample in the incomplete sample set, we denote its model input parameters as θ_j and its multi-fidelity sample snapshot as $s_j = \begin{pmatrix} s_j^L \\ s_j^H \end{pmatrix}$. Here, its low-fidelity output vector $s_j^L = (s_{1,j}^L, s_{2,j}^L, \dots, s_{n_L,j}^L)^T$ is obtained with CFD, while its high-fidelity output vector s_j^H is regarded as unknown and needs to be predicted.

The projection operator is defined as follows:

$$\Gamma = \begin{pmatrix} I_{n_L} & 0 \\ 0 & 0 \end{pmatrix}_{(n_L+n_H) \times (n_L+n_H)}$$

where I_{n_L} is the unit matrix with dimension n_L , and the expansion coefficient vector $\vec{\beta}$ of the incomplete sample snapshot in the orthogonal basis function space is obtained by solving the following extremal problem:

$$\mathcal{J}(\vec{\beta}) = \left\| \Gamma s_j - \Gamma \Phi \vec{\beta} \right\|$$

After simple matrix operations, we can obtain:

$$\vec{\beta} = (\Phi^T \Gamma \Phi)^{-1} \Phi^T \Gamma s_j$$

The $(n_L + 1)$ -th to $(n_L + n_H)$ -th elements of the vector $\Phi \vec{\beta}$ are the prediction of s_j^H .

2.6. Kriging Model

The Kriging model, a surrogate modeling approach widely used in academia and industry, is founded on the theory of random processes. It assumes the presence of spatial

correlation among the response values at different inputs and interpolates the response value at prediction location using the existing observations.

The Kriging model can be expressed as the summation of a linear regression model $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ and a stochastic process $Z(\mathbf{x})$, that is:

$$Y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x})$$

where $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})]^T$ represents regression functions, and $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_k]^T$ is the unknown regression parameters. $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ contributes a global model, and $Z(\mathbf{x})$ accounts for local deviations.

It is usually assumed that $Z(\mathbf{x})$ is a Gaussian random process with zero mean. The covariance between $Z(\mathbf{x})$ and $Z(\mathbf{w})$ is denoted as:

$$\text{Cov}(z(\mathbf{w}), z(\mathbf{x})) = \sigma^2 R(\mathbf{w}, \mathbf{x})$$

For a given experimental design, denoted as $\mathbf{S} = [s_1, \dots, s_N]$, the corresponding observation is denoted as $\mathbf{y}_s = [y(s_1), \dots, y(s_N)]$. The maximum likelihood estimates (MLEs) of the unknown parameters $\boldsymbol{\beta}$ and σ^2 , derived from these sample data, are as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y}_s \\ \hat{\sigma}^2 &= \frac{1}{N} (\mathbf{y}_s - \mathbf{F} \hat{\boldsymbol{\beta}})^T \mathbf{R}^{-1} (\mathbf{y}_s - \mathbf{F} \hat{\boldsymbol{\beta}}) \end{aligned}$$

where the regression design matrix \mathbf{F} and the correlation function matrix \mathbf{R} are, respectively, defined as:

$$\begin{aligned} \mathbf{F} &= [\mathbf{f}(s_1), \dots, \mathbf{f}(s_N)]^T \\ \mathbf{R} &= [R(s_i, s_j)]_{i,j}, \quad 1 \leq i, j \leq N \end{aligned}$$

The best linear unbiased predictor for the prediction location \mathbf{x} is given by:

$$\hat{y}(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^T(\mathbf{x})\hat{\boldsymbol{\alpha}}$$

where $\hat{\boldsymbol{\alpha}}$ is defined as $\hat{\boldsymbol{\alpha}} = \mathbf{R}^{-1}(\mathbf{y}_s - \mathbf{F}\hat{\boldsymbol{\beta}})$, and the correlation vector, denoted by \mathbf{r} , represents the correlation between the training samples \mathbf{S} and the prediction sample \mathbf{x} , that is:

$$\mathbf{r} = [R(s_1, \mathbf{x}), \dots, R(s_N, \mathbf{x})]^T$$

3. Case Description

To assess the efficacy of the proposed approach, this paper examines the influence of uncertainty in the coefficients of the SA turbulence model on the prediction of airfoil aerodynamics. The SA model is extensively utilized in aerospace and related fields [20]. The computation presumes fully turbulent flow and disregards the transition term in the original model, resulting in nine coefficients denoted as $c_{b1}, \sigma, c_{b2}, \mathcal{K}, c_{w2}, c_{w3}, c_{v1}, c_{t3}, c_{t4}$. It is assumed that the model coefficients are subject to epistemic uncertainty, which can be characterized using probabilistic methods from a credibility perspective. The model parameters are assumed to follow uniform distributions with the intervals specified in Table 1, with parameter ranges referenced from the pertinent literature [21]. It is worth noting that the mathematical representation of model uncertainty parameters, including distribution types and parameters, necessitates thorough consultation with model developers. The present study primarily focuses on the propagation of uncertainty given a mathematical description of the uncertain parameters, rather than on the precise quantification of uncertainty for the model parameters. Numerous investigations have been conducted on the quantification of uncertainty in SA turbulence model parameters, primarily employing polynomial chaos methods, with an emphasis on overall aerodynamic outputs such as airfoil or aircraft forces [21,22].

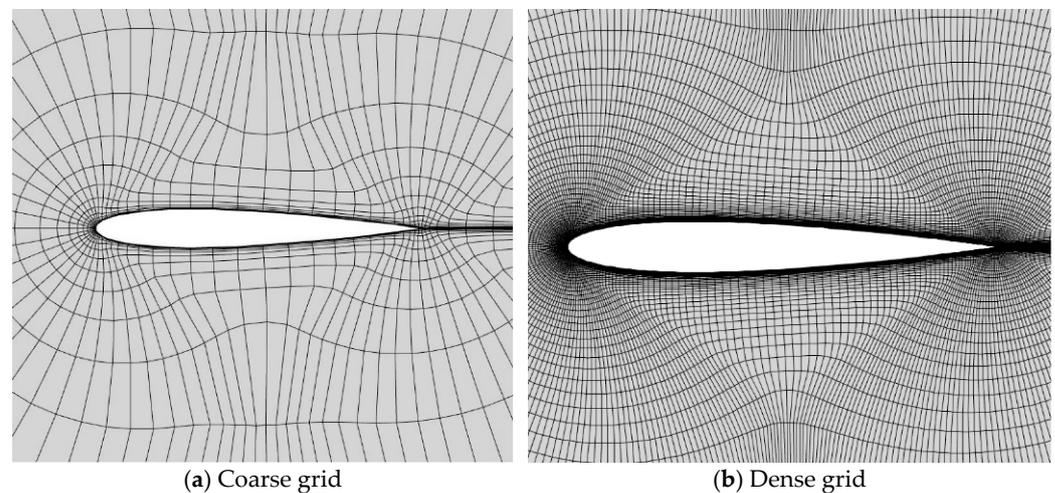
Table 1. Interval and standard value of SA model parameters.

	Minimum Value	Maximum Value	Standard Value
c_{b1}	0.12893	0.137	0.1355
σ	0.6	1.0	2/3
c_{b2}	0.60983	0.6875	0.622
κ	0.38	0.42	0.41
c_{w2}	0.055	0.3525	0.3
c_{w3}	1.75	2.5	2.0
c_{v1}	6.9	7.3	7.1
c_{t3}	1.0	2.0	1.2
c_{t4}	0.3	0.7	0.5

The two numerical examples investigated are the distribution of the wall friction coefficient in the low-speed flow over an NACA0012 airfoil and the distribution of the wall pressure coefficient in the transonic flow over an M6 wing. These two cases are presented separately in the following subsections.

3.1. Low-Speed Flow around an NACA0012 Airfoil

The NACA0012 airfoil, with a symmetric profile and 12% thickness, is considered under the computational condition of $M_\infty = 0.15$, $\alpha = 5.0^\circ$, $T_\infty = 288.15K$, $Re_\infty = 6 \times 10^6$. Two sets of grids with different densities, shown in Figure 2, are utilized for the generation of high- and low-fidelity samples. The number of cells is 3584 and 57,344, respectively. The coarse grid has a grid size that is one-sixteenth of the dense grid, and typically requires fewer iteration steps to reach convergence. However, for the purpose of uniformly assessing computational cost, it is assumed that the low-fidelity samples incur one-sixteenth of the computational cost of the high-fidelity samples.

**Figure 2.** Computational grid for NACA0012 airfoil.

The wall friction coefficient distribution obtained using standard parameters of the SA model is presented in Figure 3. The high- and low-fidelity results exhibit similar trends, characterized by local abrupt variations on the upper and lower surfaces of the airfoil leading edge, followed by a gradual stabilization. However, significant differences in magnitude are observed, particularly on the upper surface where the dense grid prediction is notably higher. Accurately capturing these local abrupt variations in friction coefficients poses a significant challenge and demands high accuracy from the prediction model.

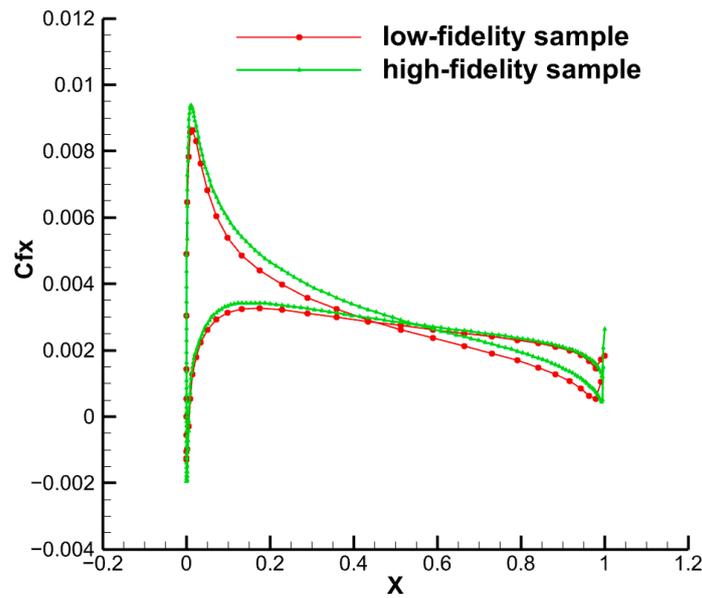


Figure 3. Wall friction coefficient distribution for NACA0012 airfoil under standard SA model parameters.

3.2. Transonic Flow around an M6 Wing

The transonic flow over the M6 wing serves as a benchmark test case for assessing transonic flow solvers. The wing features a root chord of approximately 0.8 m and a half-span of approximately 1.2 m, with a symmetric airfoil section. Two sets of grids with different densities, shown in Figure 4, are utilized for the generation of low- and high-fidelity samples. The number of cells is 990,360 and 3,594,863, respectively. Therefore, it is assumed that the acquisition cost of one high-fidelity sample is 3.5 times that of one low-fidelity sample. The output of this case constitutes a vector comprising the pressure coefficients of all grid points covering the entire wing surface, with a dimensionality of 13,638 for the coarse grid and 29,684 for the dense grid.

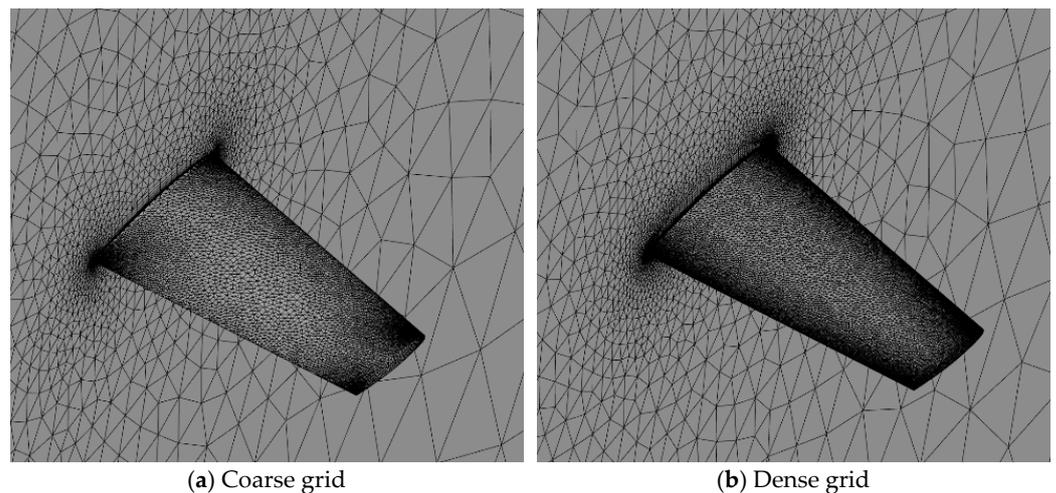


Figure 4. Computational grid for M6 wing.

The computational condition is set at $M_\infty = 0.8395$, $\alpha = 3.06^\circ$, $T_\infty = 255.56K$, $Re_\infty = 1.172 \times 10^7$. Under this condition, a λ -shaped shock structure develops on the upper surface of the wing, as shown in Figure 5, posing substantial challenges for the prediction model. The pressure distributions obtained by using two sets of grids are consistent, but

near the suction peak and shock wave on the upper surface, the dense grid results are in better agreement with the reference experiment results shown in Figure 6.

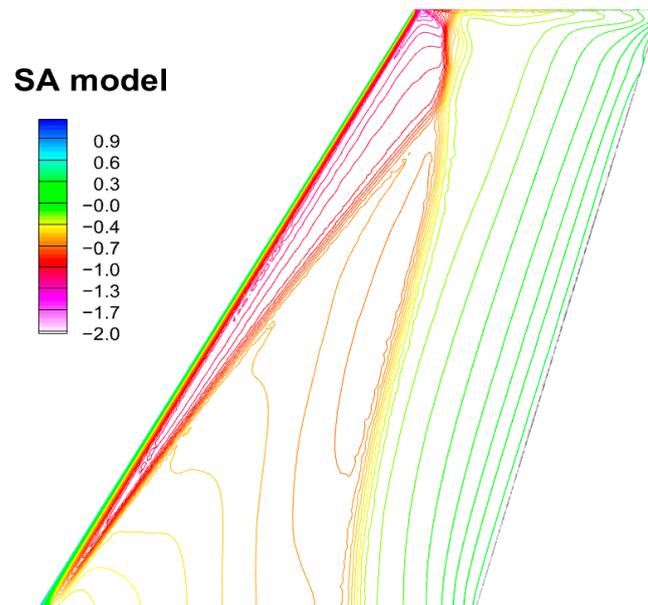


Figure 5. Pressure contours for M6 wing under standard model parameters using fine grid.

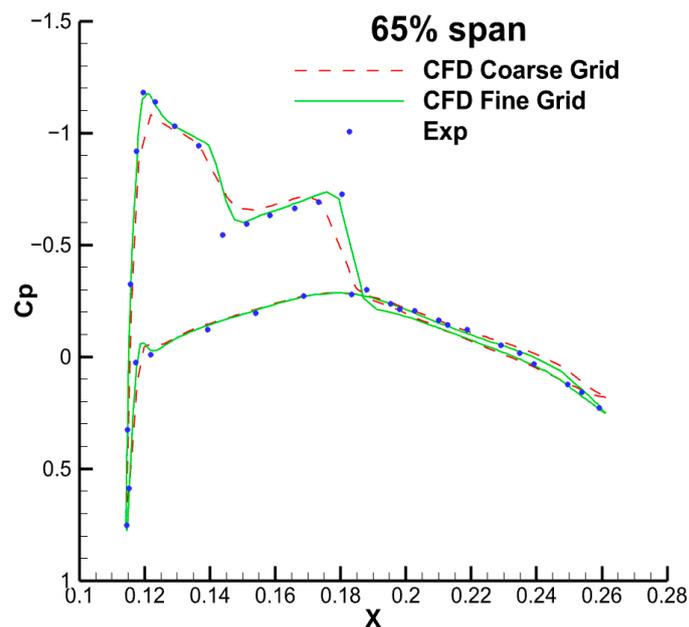


Figure 6. Pressure coefficient distribution for M6 wing under original SA model parameter at 65% wingspan.

The sample data in this study are generated using the in-house unstructured grid solver Flowstar [23], which is founded on a cell-centered finite volume methodology and is adept at handling diverse element types, including hexahedra, tetrahedra, prisms, pyramids, and other polyhedra generated via geometrical multigrid techniques. Second-order accuracy in space is attained through linear reconstruction within cells. The vertex-based Green–Gauss approach [24] is employed for gradient computations to uphold accuracy and robustness. To mitigate oscillations in regions of high gradients, Venkatakrishnan’s limiter [25] is utilized. The Roe scheme is engaged for inviscid flux computations. A first-order backward Euler time-differencing scheme with local time stepping is implemented

to approximate a steady state, facilitating convergence. The flux Jacobian is derived from a first-order upwind scheme, with the divided convective flux Jacobian composed of the convective flux Jacobian and its spectral radius. The viscous flux Jacobian is approximated using its spectral radius.

In Kriging modeling, the regression function is a constant function, and the covariance function, which determines the spatial correlation structure, is a Gaussian function. In the POD method, the generalized energy criterion is set at 99.9%.

4. Results and Discussion

The prediction accuracy of the model can be assessed through the prediction error of the high-fidelity output vector s^H . For incomplete samples within the training dataset, the error vector is defined as $\vec{\varepsilon} = s^H|_{\text{Gappy POD}} - s^H|_{\text{CFD}}$, serving as a metric for evaluating the accuracy of the Gappy POD method. For the testing sample, the error vector is defined as $\vec{\varepsilon} = s^H|_{\text{model}} - s^H|_{\text{CFD}}$, serving to evaluate the accuracy of the overall model.

For each sample, we define dimensionless errors based on the 1-norm, 2-norm, and infinity norm of the error vector:

$$\varepsilon_1 = \frac{\|\vec{\varepsilon}\|_1}{n_H \cdot \text{ref}}, \quad \varepsilon_2 = \frac{\|\vec{\varepsilon}\|_2}{\sqrt{n_H} \cdot \text{ref}}, \quad \varepsilon_{\text{inf}} = \frac{\|\vec{\varepsilon}\|_{\text{inf}}}{\text{ref}}$$

where ref is the range of the sample response, $\text{ref} = \max(s^H|_{\text{CFD}}) - \min(s^H|_{\text{CFD}})$.

4.1. Low-Speed Flow around an NACA0012 Airfoil

In this case, we fix $n_{\text{train}} = 32$ and $n_{\text{test}} = 49$ and assign M values of 11, 13, 15, and 17. This configuration facilitates the analysis of the cost and accuracy trade-offs of the overall model in this paper relative to modeling with high-fidelity samples only. Specifically, the computational cost of 32 low-fidelity samples is deemed equivalent to that of 2 high-fidelity samples.

(1) Accuracy analysis of Gappy POD method

Given that high-fidelity sample data serve as the basis for subsequent POD and Kriging modeling, it is crucial to first analyze the accuracy of the Gappy POD method in reconstructing the corresponding high-fidelity output from low-fidelity output. In the analysis, M is set to 11, with similar results observed for M values of 13, 15, and 17.

Randomly selecting an incomplete sample from the training dataset, Figure 7 presents a comparison between the wall friction coefficient distribution predicted using the Gappy POD method and high-fidelity CFD computation. Visually, the disparity between the two is minimal.

To further evaluate the accuracy of the Gappy POD method, the prediction error was computed for all incomplete samples; the resulting error distribution is depicted in Figure 8. Across all incomplete samples, it can be noted that ε_1 remains below 0.25%, ε_2 remains below 0.6%, and ε_{inf} remains below 2.5%. Considering the randomness of the Latin hypercube sampling and exchange algorithm, steps (a) through (d) in the training process were repeated a total of 10 times. Figure 9 presents the statistical analysis of errors for all incomplete samples, revealing median values of 0.0007, 0.0015, and 0.0070 for ε_1 , ε_2 , and ε_{inf} , respectively. The results demonstrate that the Gappy POD approach is capable of reconstructing high-fidelity outputs using a limited amount of complete sample data and low-fidelity outputs, effectively substituting for computationally expensive high-fidelity CFD calculations and significantly reducing computational costs.

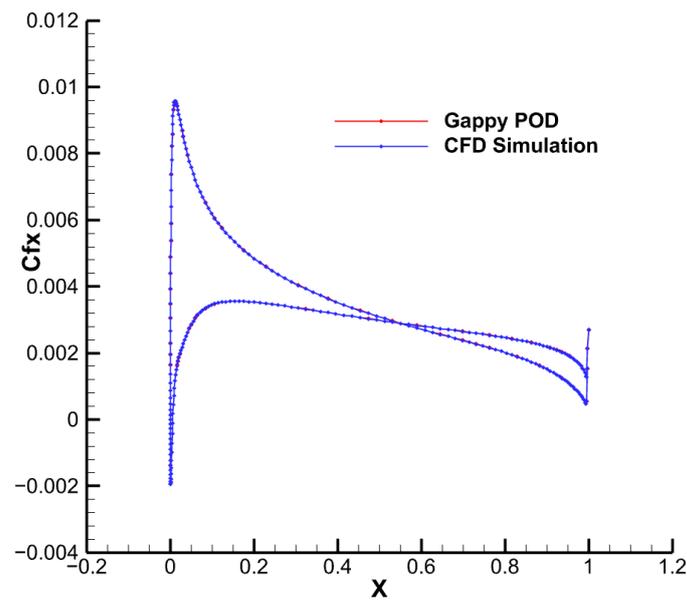


Figure 7. Comparison between the wall friction coefficients predicted using the Gappy POD method and high-fidelity CFD computation for the NACA0012 case.

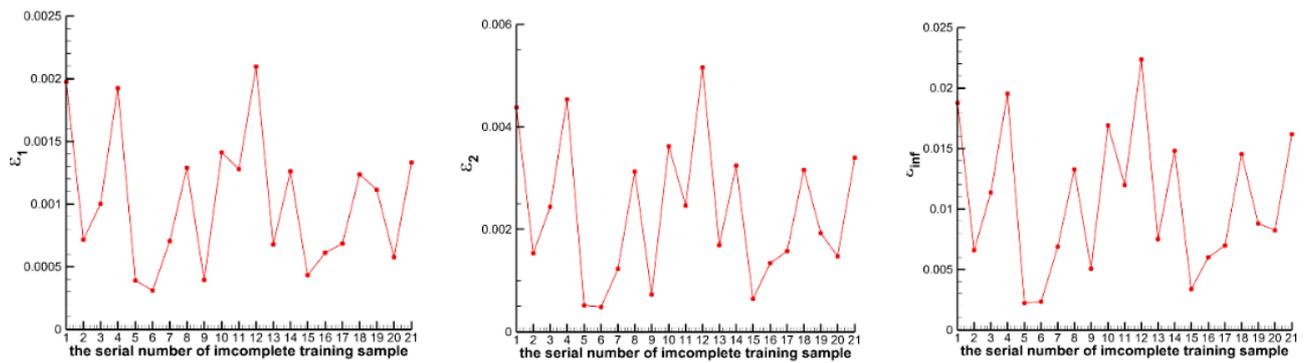


Figure 8. Error analysis of Gappy POD method for predictions on incomplete samples for the NACA0012 case.

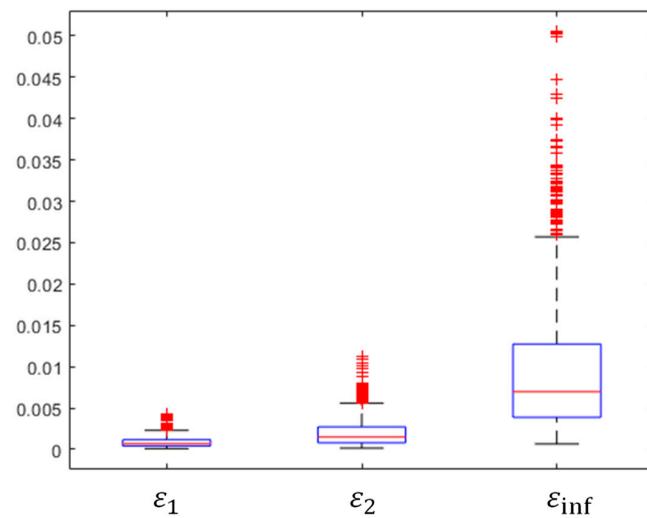


Figure 9. Statistical error analysis of Gappy POD method for predictions on incomplete samples (repeated 10 times).

(2) Accuracy analysis of overall model

Given that this study employs the Monte Carlo method to perform extensive random sampling on the overall model, it is imperative to validate the predictive capability of the model. Figure 10 presents a comparison between the wall friction coefficient distribution predicted by the overall model and the high-fidelity CFD computation for a randomly selected sample from the test dataset with $M = 11$. Visually, the two exhibit excellent agreement.

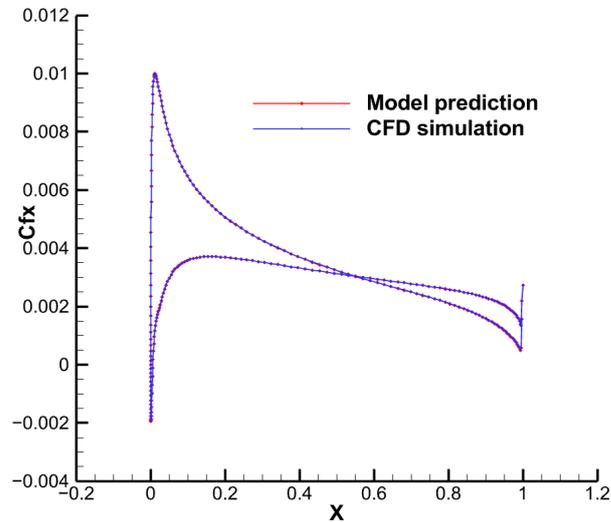


Figure 10. Comparison between the wall friction coefficients predicted by the overall model and the high-fidelity CFD computation.

The prediction error was computed for all testing samples, and the resulting error distribution is depicted in Figure 11. Across all testing samples, it can be noted that ε_1 remains below 0.25%, ε_2 remains below 0.5%, and ε_{inf} remains below 2.5%. The results strongly validate the predictive capability of the proposed model in accurately estimating the wall friction coefficient distribution for new samples, thus providing robust support for large-scale random sampling processes.

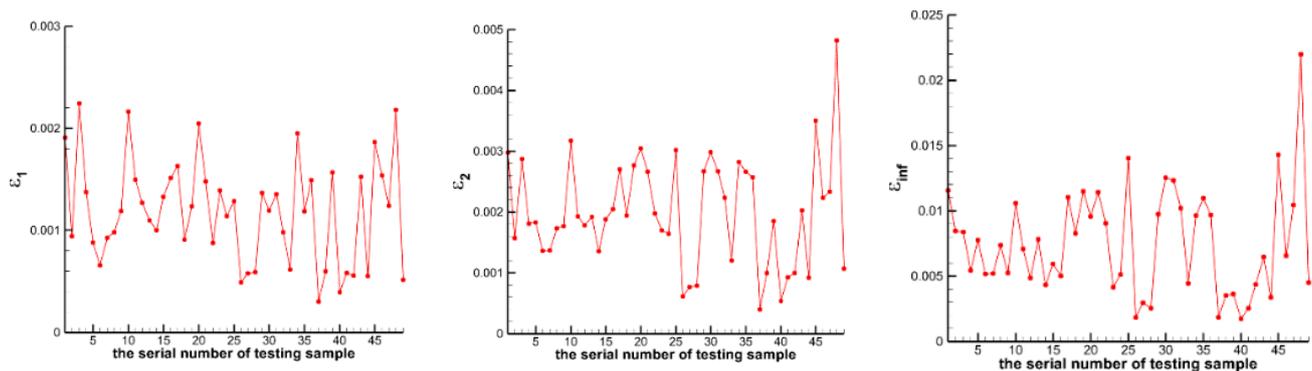


Figure 11. Error analysis of overall model for predictions on testing samples for the NAC0012 case.

(3) Analysis of the influence of sample size

In the context of multidimensional correlated responses, while the multi-fidelity modeling approach presented in this study is applicable, an alternative approach could involve utilizing solely high-fidelity data for POD decomposition and surrogate model construction. Given that low-fidelity samples, despite their lower acquisition costs, still require CFD computations, a rigorous examination is warranted to determine the precise impact of incorporating additional low-fidelity calculations on the overall modeling accuracy.

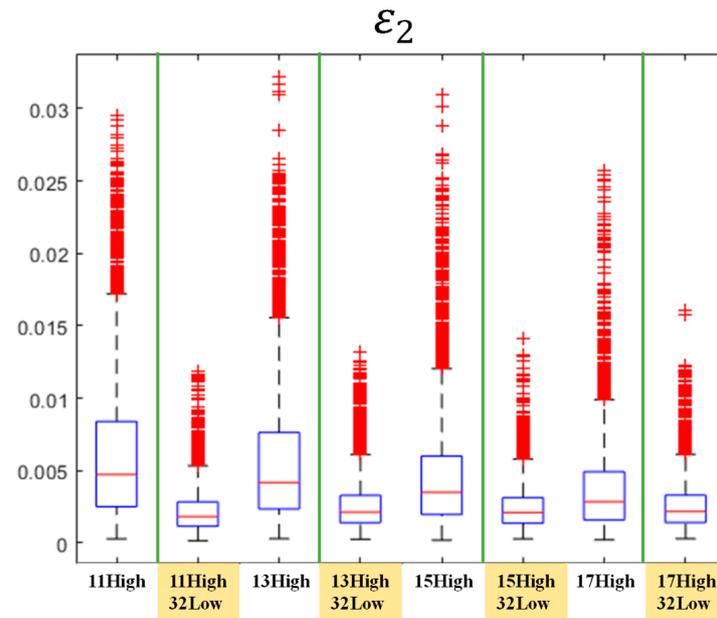


Figure 13. Statistical error analysis for predictions on testing samples (ϵ_2).

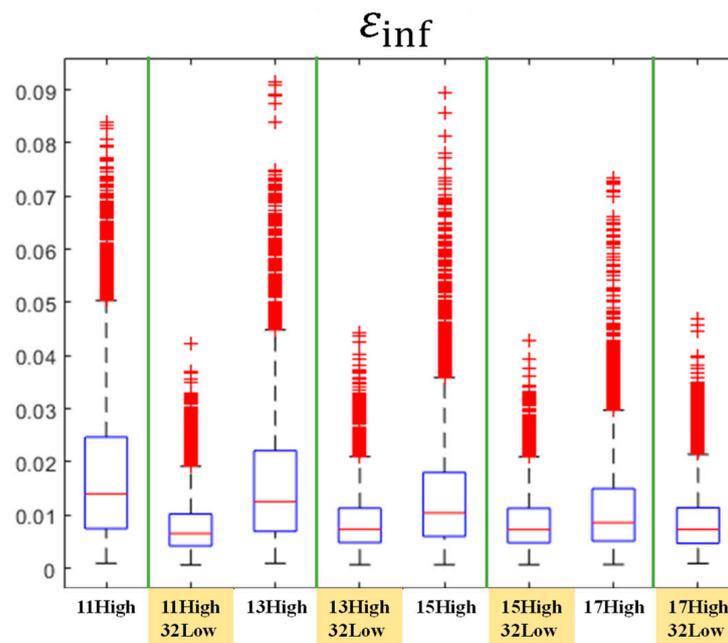


Figure 14. Statistical error analysis for predictions on testing samples (ϵ_{inf}).

The relative significance of the nine model parameters on the distribution of the friction coefficients can be ascertained through the sensitivity analysis approach, as presented in Table 2. A global sensitivity analysis approach proposed by Gamboa et al., which decomposes the covariance of multi-output variables [26], is utilized. As shown, κ emerges as the most influential parameter, which can be expected from a physical standpoint. In the constant Reynolds stress layer of plane shear turbulence, the eddy viscosity coefficient is directly proportional to κ , i.e., $\nu_t = \kappa u_\tau d$. The eddy viscosity coefficient characterizes the relationship between Reynolds stress and mean flow, serving a role analogous to molecular viscosity in RANS simulations. Hence, κ is intimately linked with the viscosity of the flow, and the friction coefficient is directly correlated with the flow viscosity.

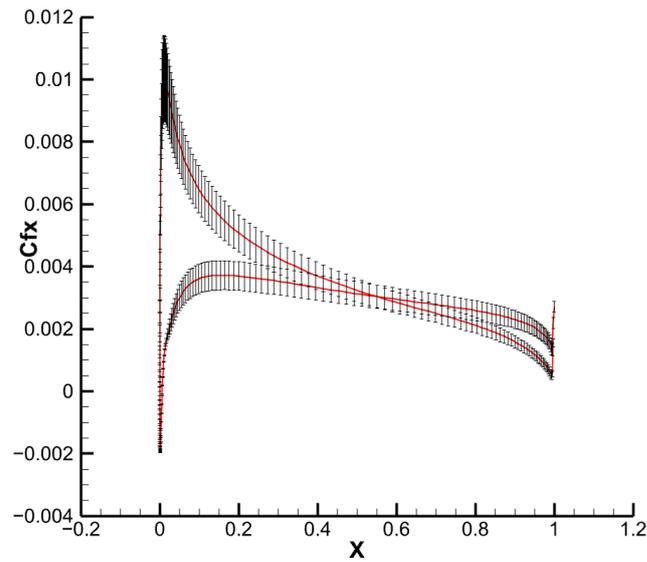


Figure 15. Mean value and 99% confidence interval of the wall friction coefficient.

Table 2. Sobol indicators of SA model parameters.

	Main Effect	Total Effect
c_{b1}	0.0061	0.0073
σ	0.1223	0.1270
c_{b2}	0.0001	0.0004
κ	0.7945	0.7991
c_{w2}	0.0482	0.0504
c_{w3}	0.0001	0.0001
c_{v1}	0.0200	0.0210
c_{t3}	0.0001	0.0008
c_{t4}	0.0001	0.0004

4.2. Transonic Flow around an M6 Wing

(1) Accuracy analysis of Gappy POD method

In the analysis, n_{train} is set to 21 and M is 6. Randomly selecting an incomplete sample from the training dataset, Figure 16 presents a comparison between the pressure coefficient distribution predicted by the Gappy POD method and the high-fidelity CFD computation. Visually, the disparity between the two is minimal.

To further evaluate the accuracy of the Gappy POD method, the prediction error was computed for all incomplete samples, and the resulting error distribution is depicted in Figure 17. Across all incomplete samples, it can be noted that ε_1 remains below 0.0035%, ε_2 remains below 0.03%, and ε_{inf} remains below 1.5%. The results demonstrate that the Gappy POD approach is capable of reconstructing high-fidelity outputs using a limited amount of complete sample data and low-fidelity outputs, effectively substituting for computationally expensive high-fidelity CFD calculations and significantly reducing computational costs.

(2) Accuracy analysis of overall model

The overall model was constructed using n_{train} high-fidelity responses, of which M responses were calculated using CFD simulations, and the rest were reconstructed using the Gappy POD algorithm from its low-fidelity responses. Figure 18 presents a comparison between the pressure coefficient predicted by the overall model and the high-fidelity CFD computation for a randomly selected sample from the test dataset. Visually, the two exhibit excellent agreement.

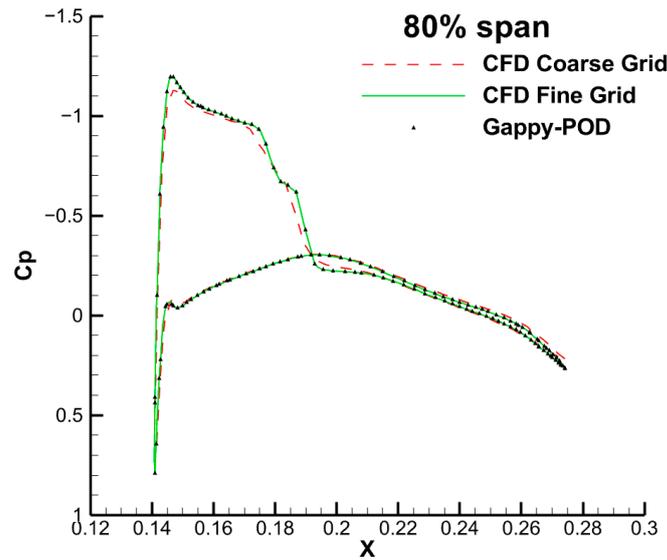


Figure 16. Comparison between the pressure coefficients predicted using the Gappy POD method and the high-fidelity CFD computation for the M6 case.

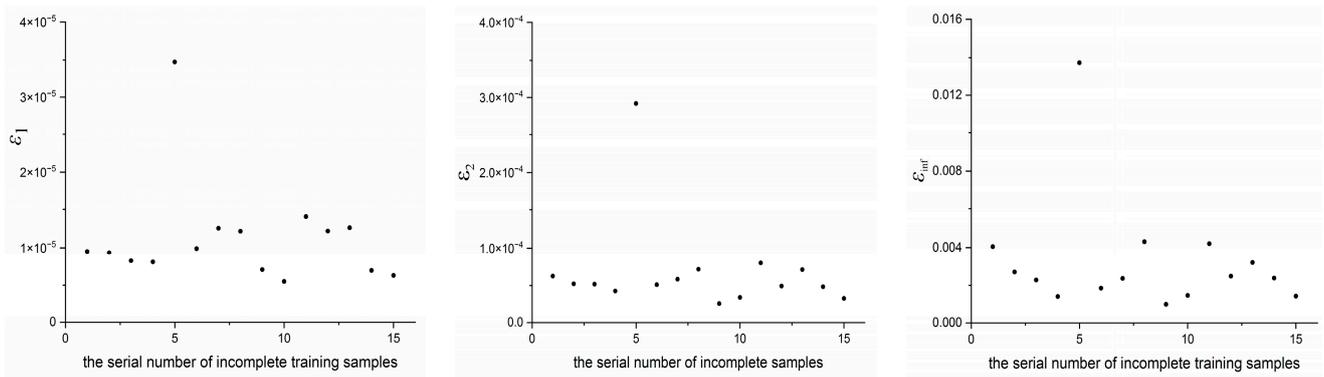


Figure 17. Error analysis of Gappy POD method for predictions on incomplete samples for the M6 case.

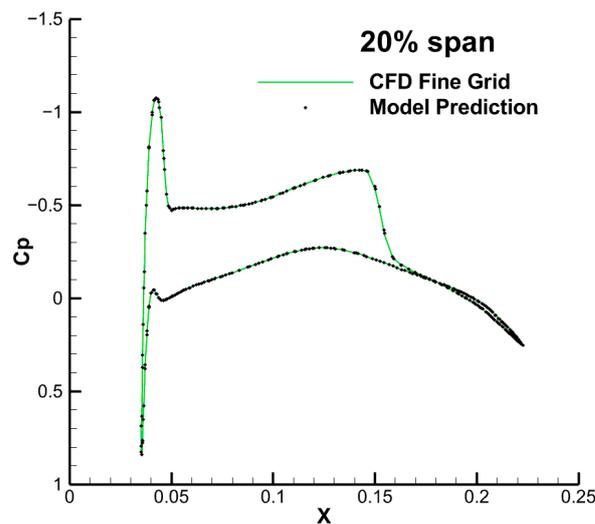


Figure 18. Comparison between the pressure coefficients predicted by the overall model and the high-fidelity CFD computation for the M6 case.

The prediction error was computed for all testing samples, and the resulting error distribution is depicted in Figure 19. Across all testing samples, it can be noted that ε_1 remains below 0.009%, ε_2 remains below 0.035%, and ε_{inf} remains below 2%. The results strongly validate the predictive capability of the proposed model in accurately estimating the pressure coefficient distribution for new samples, thus providing robust support for large-scale random sampling processes.

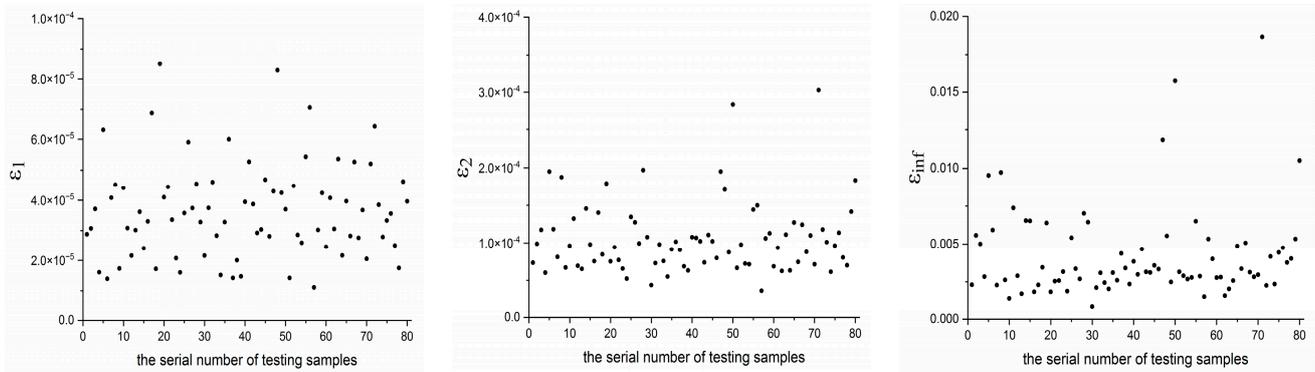


Figure 19. Error analysis of overall model for predictions on testing samples for the M6 case.

(3) Analysis of the influence of sample size

To demonstrate the advantages of the proposed multi-fidelity model over the single-fidelity model using only high-fidelity samples, we analyzed the prediction performance of the two approaches. The computational cost of the low-fidelity sample is converted into the equivalent cost of the high-fidelity sample, with a ratio of 1 to 3.5. The mean prediction error of the two approaches is shown in Figure 20, demonstrating that with the inclusion of low-fidelity samples, the multi-fidelity model can achieve a much lower prediction error with approximately the same computational cost compared with the single-fidelity model.

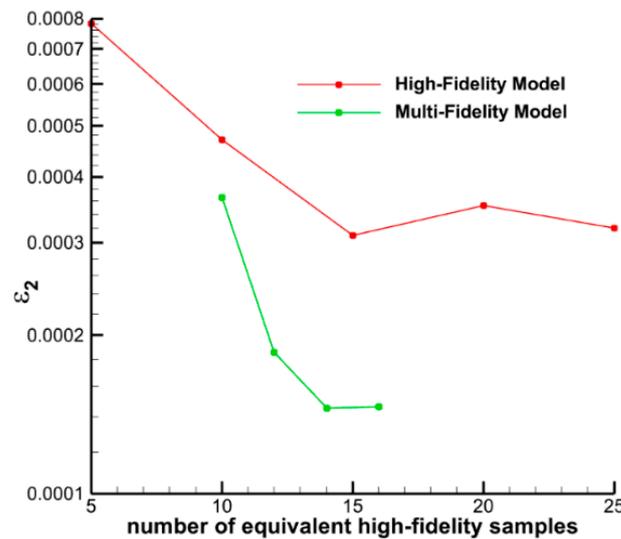


Figure 20. Comparison of prediction error on testing samples between multi-fidelity and high-fidelity models for the M6 case.

(4) Statistical results analysis

Latin hypercube sampling was employed again within the nine-dimensional input space to generate 10^6 samples. These inputs were sequentially fed into the fast prediction model to acquire the corresponding pressure coefficients. The statistical analysis of the sample data yielded the mean value and standard deviation of the wall pressure coefficient, as depicted in Figure 21. Notably, the uncertainty of the pressure coefficient is concentrated

at the λ -shaped shock on the upper surface, whereas the uncertainty at other locations is minimal. This observation aligns with expectations, given that the prediction of shocks is highly sensitive to turbulence model parameters.

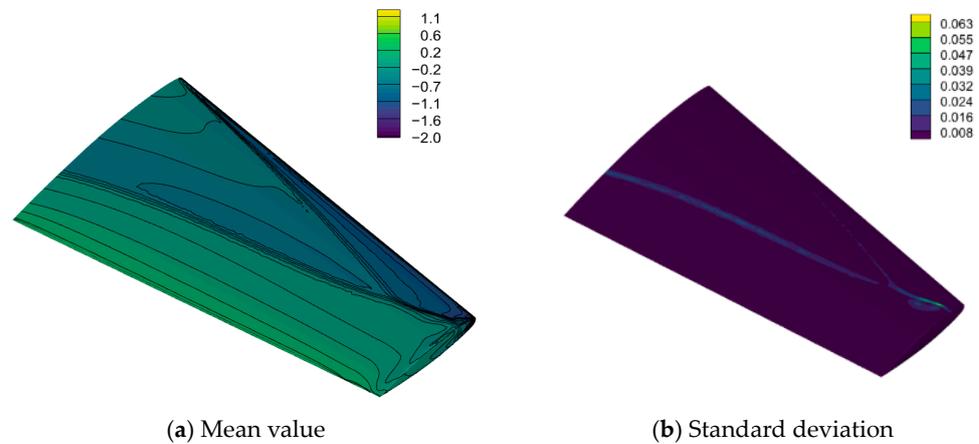


Figure 21. Contours of mean value and standard deviation of the wall pressure coefficients.

Through the application of the multi-output global sensitivity analysis approach grounded in covariance decomposition, the relative significance of model parameters on the pressure coefficient distribution can be ascertained, as presented in Table 3. Again, κ emerges as the most influential parameter. Following the same analysis as the first case, κ is closely linked with the viscosity of the flow, which significantly impacts the kinetic energy loss of fluid micro-groups within the boundary layer and the capacity to resist adverse pressure gradients, ultimately influencing the prediction of shock wave positions.

Table 3. Sobol indicators of SA model parameters for M6 case.

	Main Effect	Total Effect
c_{b1}	0.0625	0.0674
σ	0.2375	0.2550
c_{b2}	0.0050	0.0050
κ	0.6079	0.6184
c_{w2}	0.0338	0.0382
c_{w3}	0.0016	0.0017
c_{v1}	0.0340	0.0368
c_{t3}	0.0035	0.0040
c_{t4}	0.0012	0.0012

5. Conclusions

This study addresses the requirement for uncertainty quantification in multidimensional correlated responses within flow fields. Building on the previously established modeling framework, which utilizes Proper Orthogonal Decomposition for flow field reduction and surrogate models, we introduce a multi-fidelity modeling framework that integrates high- and low-fidelity sample data. The findings of this investigation demonstrate the following:

- (1) The Gappy POD-based method for supplementing missing data in flow fields enables the restoration of high-fidelity outputs from a limited amount of complete sample data, utilizing the low-fidelity outputs of incomplete samples. This approach effectively avoids the need for computationally expensive high-fidelity CFD calculations on a large number of samples, significantly reducing computational costs.
- (2) The multi-fidelity modeling approach demonstrates a marked improvement in prediction accuracy and model stability compared to single-fidelity methods while incurring

approximately the same computational cost for sample processing. This methodology offers an efficient and robust prediction model for large-scale random sampling.

The cases investigated in this study demonstrate a relatively stable variation in the flow field under differing turbulence model coefficients. In such cases, it is commonly accepted that Proper Orthogonal Decomposition effectively captures the fundamental modes of the flow field. However, if the flow field exhibits rapid changes in response to model parameters, such as smooth flow fields under certain parameters and discontinuous flow fields under others, further validation is required to ascertain the accuracy of the developed model in predicting flow field variables. Autoencoder and deep neural network approaches represent potential solutions to these challenges and will be the focus of future investigations.

Author Contributions: Methodology, W.X. and J.C.; Validation, L.L. and X.W.; Investigation, W.Z.; Software, J.Z.; Writing—original draft, W.X., L.L. and J.Z.; Writing—review & editing, J.C., W.Z. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NSAF (grant no. U2230208) and the National Numerical Wind Tunnel Project. The APC was funded by NSAF (grant no. U2230208).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mehta, U.B.; Eklund, D.R.; Romero, V.J.; Pearce, J.A.; Keim, N.S. *Simulation Credibility, Advances in Verification, Validation, and Uncertainty Quantification*; NASA/TP-2016-219422; NASA Ames Research Center: Moffett Field, CA, USA, 2016.
2. Xiu, D.; Karniadakis, G. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comput.* **2002**, *24*, 619–644. [[CrossRef](#)]
3. Li, M.; Wang, Z. Surrogate Model Uncertainty Quantification for Reliability-based Design Optimization. *Reliab. Eng. Syst. Saf.* **2019**, *192*, 106432. [[CrossRef](#)]
4. Bhattacharyya, B. Uncertainty quantification of dynamical systems by a POD–Kriging surrogate model. *J. Comput. Sci.* **2022**, *60*, 101602. [[CrossRef](#)]
5. Li, X.; Zou, Z.-J.; Wang, Z.-H.; Zou, L.; Gao, H. Surrogate model based uncertainty quantification of CFD simulations of the viscous flow around a ship advancing in shallow water. *Ocean. Eng.* **2021**, *234*, 109206.
6. Tripathy, R.K.; Bilionis, I. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *J. Comput. Phys.* **2018**, *375*, 565–588. [[CrossRef](#)]
7. Kennedy, M.C.; O’Hagan, A. Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. *Biometrika* **2000**, *87*, 1–13. [[CrossRef](#)]
8. Forrester, A.I.J.; Sóbester, A.; Keane, A.J. Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc. A* **2007**, *463*, 3251–3269. [[CrossRef](#)]
9. Zimmermann, R.; Han, Z.H. Simplified cross-correlation estimation for multifidelity surrogate cokriging models. *Adv. Appl. Math. Sci.* **2010**, *7*, 181–202.
10. Meng, X.; Karniadakis, G.E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *J. Comput. Phys.* **2019**, *401*, 109020. [[CrossRef](#)]
11. Motamed, M. A multi-fidelity neural network surrogate sampling method for uncertainty quantification. *J. Comput. Phys.* **2021**, *426*, 109923. [[CrossRef](#)]
12. Cook, R.D.; Nachtsheim, C.J. A comparison of algorithms for constructing exact D-optimal designs. *Technometrics* **1980**, *22*, 315–324. [[CrossRef](#)]
13. Morris, M.D.; Mitchell, T.J. Exploratory designs for computer experiments. *J. Stat. Plan. Inference* **1995**, *43*, 381–402. [[CrossRef](#)]
14. Everson, R.; Sirovich, L. Karhunen-Loeve procedure for gappy data. *J. Opt. Soc. Am. A* **1995**, *12*, 1657–1664. [[CrossRef](#)]
15. Benamara, T.; Breitkopf, P.; Lepot, I.; Sainvitu, C. Adaptive infill sampling criterion for multi-fidelity optimization based on Gappy-POD Application to the flight domain study of a transonic airfoil. *Struct. Multidiscip. Optim.* **2016**, *54*, 843–855. [[CrossRef](#)]
16. Lumley, J.L. The Structure of Inhomogeneous Turbulent Flows. *Commun. Pure Appl. Math.* **1967**, *20*, 453–488. [[CrossRef](#)]
17. Sirovich, L.; Kirby, M. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* **1987**, *4*, 519–524. [[CrossRef](#)] [[PubMed](#)]
18. Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem. Metall. Min. Eng. Soc. S. Afr.* **1951**, *52*, 119–139.

19. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and analysis of computer experiments. *Stat. Sci.* **1989**, *4*, 409–423. [[CrossRef](#)]
20. Spalart, P.R.; Allmaras, S.R. A One-Equation Turbulence Model for Aerodynamic Flows. *AIAA J.* **1992**, *30*, 5–12. [[CrossRef](#)]
21. Schaefer, J.; Cary, A.; Mani, M.; Spalart, P. Uncertainty Quantification and Sensitivity Analysis of SA Turbulence Model Coefficients in Two and Three Dimensions. In Proceedings of the 55th AIAA Aerospace Sciences Meeting, Grapevine, TX, USA, 9–13 January 2017. AIAA Paper 2017-1710.
22. Stephanopoulos, K.; Witte, I.; Wray, T.J.; Agarwal, R.K. Uncertainty Quantification of Turbulence Model Coefficients in OpenFOAM and Fluent for Mildly Separated Flows. In Proceedings of the 46th AIAA Fluid Dynamics Conference, Washington, DC, USA, 13–17 June 2016. AIAA Paper 2016-4401.
23. Chen, J.Q.; Wu, X.J.; Zhang, J.; Li, B.; Jia, H.Y.; Zhou, N.C. Flowstar: General Unstructured-grid CFD Software for National Numerical Wind Tunnel (NNW) Project. *Acta Aeronaut. Astronaut. Sin.* **2021**, *42*, 625739. (In Chinese)
24. Diskin, B.; Thomas, J.L. Comparison of Node-Centered and Cell-Centered Unstructured Finite Volume Discretizations: Inviscid Fluxes. *AIAA J.* **2011**, *49*, 836–854. [[CrossRef](#)]
25. Venkatakrisnan, V. On the Accuracy of Limiters and Convergence to Steady-State Solutions. In Proceedings of the 31st Aerospace Sciences Meeting, Reno, NV, USA, 11–14 January 1993. AIAA Paper 1993-0880.
26. Gamboa, F.; Janon, A.; Klein, T.; Lagnoux, A. Sensitivity Analysis for Multidimensional and Functional Outputs. *Electron. J. Stat.* **2014**, *8*, 575–603. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.