A high-fidelity polynomial chaos modified method suitable for CFD uncertainty quantification

Jiangtao Chen 1,2, Chao Zhang 1,2, Wei Zhao 1,* and Xiaojun Wu 1

1 China Aerodynamics Research and Development Center, Mianyang, China
2 State Key Laboratory of Aerodynamics, Mianyang, China

*Corresponding author e-mail: weizhao@cardc.cn

Abstract. There are many sources of uncertainty in the process of numerical simulation for engineering configure. Evaluating and quantifying the uncertainty of simulation results is very important for the design and evaluation process of the industrial department. The non-intrusive polynomial chaos method is a commonly used uncertainty quantification method, but a polynomial chaos with high accuracy requires a sufficient number of high-fidelity samples, which causes "dimension disaster" for high-dimensional problems. This paper proposes a modified method. First, a low-fidelity calculation model is used to generate a low-fidelity polynomial chaos, and then a small amount of high-fidelity calculation data is used to modify the low-fidelity polynomial chaos. This method is used to analyze the parameter uncertainty of the SA model in the RAE2822 airfoil calculation. Under the premise of effectively ensuring the accuracy, the calculation time is reduced by 60% compared with the method of using the full high-fidelity calculation model.

1. Introduction

CFD (Computational Fluid Dynamics) has played an increasingly important role in aerospace, land and water transportation, energy and power, atmosphere and ocean. However, there are a lot of uncertain parameters in CFD, such as turbulence model coefficients, thermodynamic parameters, etc. This also leads to significant uncertainty in the simulation results. Uncertainty factors may cause product performance fluctuations or even functional failure. NASA has investigated 2500 vehicle failures in orbit, of which approximately 52% are caused by uncertainties. Therefore, in the process of aircraft optimization design and performance evaluation, it is necessary to quantify the influence of parameter uncertainty on numerical simulation.

There are many uncertain parameters in CFD, showing significant high-dimensional characteristics. The polynomial chaos (PC) method [1] is the most commonly used method to quantify parameter uncertainties. With the increase in the dimension of uncertainty parameters, this method requires a sharp increase in the amount of calculation, which limits its application in engineering problems. This also makes it difficult for the industrial sector to evaluate the potential of many uncertain parameters on product performance. Therefore, there is an urgent need to develop more efficient methods to provide practical solutions to the problem of parameter uncertainty quantification in engineering under limited computing resources.
2. A high-fidelity polynomial chaos modified method

2.1. Polynomial Chaos

Take the single input variable as an example to introduce the PC method [2]. Assume that the random input variable \( \xi \) satisfies a certain probability density distribution \( f(\xi) \). For any random output variable \( y \), it can be expanded in the spectral space formed by the orthogonal basis function sequence of the input variable \( \xi \), where the expression of the p-order expansion is,

\[
y = \sum_{j=0}^{p} \alpha_j \psi_j(\xi)
\]

Among them \( \alpha_j \) is the deterministic component mentioned above, which can also be called degrees of freedom, and \( \psi_j \) is the basis function of random variables. The choice of the orthogonal basis function is determined by the probability density function of the random input variable, satisfying:

\[
\langle \psi_i \psi_j \rangle = \int \psi_i(\xi) \psi_j(\xi) f(\xi) d\xi = \delta_{ij} \langle \psi_i \psi_i \rangle
\]

That is, \( \{\psi_j, j \geq 0\} \) is an orthogonal polynomial sequence with a weight function of \( f(\xi) \). For uniformly distributed random variables, the optimal basis function sequence is Legendre polynomial. For normally distributed random variables, the optimal basis function sequence is Hermite polynomial. The optimality mentioned here refers to that the rate of convergence of the statistical information output by the system as the polynomial order increases is consistent with the theory.

The non-intrusive polynomial chaos method does not need to modify the control equation, and treats the existing solver as a black box. First, obtain a series of sampling points in the random space of the input variables through appropriate sampling methods, and calculate the output response at sampling point. Secondly, the number of sampling points is greater than the number of degrees of freedom in the PC expansion. So an overdetermined system of equations is formed,

\[
\Psi \hat{\alpha} = Y
\]

Where \( \Psi \) is the measurement matrix, \( \psi_j = \psi_j(\xi) \). The vector \( \hat{\alpha} \) of degrees of freedom can be obtained by the least square method. The number of sampling points \( N \) is related to the number of degrees of freedom in PC expansion, namely:

\[
N = n_p (p+1) = n_p \frac{(n + p)!}{n! p!}
\]

Where \( n \) is the dimension of the input random variable, \( p \) is the order of the polynomial chaos, and \( n_p \) is defined as the oversampling rate and is taken as 2 in the calculations in this paper. Finally, statistical information such as the output mean \( \mu \) and standard deviation \( \sigma \) can be directly obtained according to the orthogonal characteristic of the basis function.

\[
\mu(y) = \alpha_0 \\
\sigma^2(y) = \sum_{j=1}^{p} \alpha_j^2 \langle \psi_j \psi_j \rangle
\]
random input parameters to the output response in the entire value range and the interaction of random input parameters.

2.2. A high-fidelity modified method

In the multi-fidelity model, the final high-fidelity model is a combination of the low-fidelity model and the correction term. The correction term can be taken as the addition term, product term or the combination based on the low-fidelity model. The modified method is divided into the following three steps,

1) At the low-fidelity model level, a low-fidelity PC model is obtained through a few calculations.
2) Perform a small amount of deterministic calculations at the high-fidelity model level.
3) Modify the low-fidelity model established before to get the final high-fidelity model.

Here we only consider the addition term, namely:

\[ R_{\text{high}}(\xi) = R_{\text{low}}(\xi) + C(\xi) \]  \hspace{1cm} (6)

Among them, \( R_{\text{high}}(\xi) \) is the final high-fidelity PC expansion, \( R_{\text{low}}(\xi) \) is the low-fidelity PC expansion, and \( C(\xi) \) is the correction term.

Since the correction items need to be calculated with high fidelity, in order to save the calculation cost, it is necessary to determine which terms of the low fidelity PC should be corrected. In this paper, first calculate the contribution of each expansion item to variance based on the low-fidelity model, and then sort it, select several expansion items to modify its degree of freedom according to the contribution. As for the selection of several items for correction, which involves model evaluation and selection, this article is determined through the cross-validation method. The specific methods are as follows:

First, all high-fidelity sample points are randomly divided into \( k \) parts, and \( k-1 \) of them are used as learning or training samples each time to train prediction models. The remaining one is a validation sample to evaluate the generalization error of the trained model. In this way, \( k \) sets of training or testing samples are obtained. Define the \( R^2 \) coefficient of determination on the validation sample to evaluate the accuracy of the model,

\[ R^2 = 1 - \frac{\sum_{i=0}^{n_{\text{sample}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{sample}}-1} (y_i - \bar{y})^2} \]  \hspace{1cm} (7)

Where \( y_i \) is the system response calculated by the determined CFD, \( \bar{y} \) is the mean of the response, \( \hat{y}_i \) is the predicted value obtained from the model, and \( n_{\text{sample}} \) is the number of verification sample points.

Next, the accuracy of the prediction model is evaluated by the mean of the \( R^2 \) coefficient of determination on the \( k \) test set, and the variance is used to evaluate the stability of the prediction model.

Finally, choose a model with both higher accuracy and smaller variance, and relearn on all high-fidelity sample points to obtain the final high-fidelity model.

3. Numerical case

The calculation example considered in this paper is the RAE2822 airfoil [4]. This is a typical supercritical airfoil and a classic case for testing the transonic flow simulation capability of the CFD program. The calculation state of this paper is: \( Ma=0.729 \), \( Rec=6.5\times106 \), \( \alpha=2.31^\circ \). The grid used is shown in Figure 1.
The calculation program is MFlow [5] and the flow is assumed fully turbulent with ignoring the trip term in the original model. The parameters $c_{b1}, \sigma, c_{b2}, K, c_{w2}, c_{w3}, c_{t1}, c_{t3}, c_{t4}$ in the SA model, as shown in Table 1, are uncertain.

Before quantifying the uncertainty, it should be pointed out that this paper focuses on the influence of the uncertainty of the SA model coefficients, without considering the uncertainty brought by the grid generation, numerical format, etc.

The so-called high-fidelity model and low-fidelity model in this paper are distinguished by computational grid. The grid amount of the low-fidelity model is only 1/4 of the original grid, and the calculation time on this grid is also greatly reduced. Figure 2 shows the lift and drag coefficients of all sampling points calculated by the two sets of grids under the same model input parameters. It can be seen that the results of the two sets of grids are significantly different. Table 2 shows the mean and mean square deviation of the lift and drag coefficients obtained by the low-fidelity model, the directly established high-fidelity model and the modified model. There is a big difference between the low-fidelity and high-fidelity model, but little difference between the high-fidelity and modified model. In general, the finer grid calculation results are more reliable, so it can be said that the PC expansion based on the coarse grid is difficult to meet the accuracy requirements, and it must be corrected.

Table 3 shows the number of deterministic CFD calculations performed by the two high-fidelity models. The calculation time on the coarse grid is about 20% of the calculation time on the original grid. Therefore, the calculation time required to establish a high-fidelity model through the modified method is about 40% of the original method, which greatly saves the calculation cost.

### Table 1. SA model closure coefficients and associated support

| Parameter | Minimum | Maximum | Original Value |
|-----------|---------|---------|---------------|
| $c_{b1}$  | 0.12893 | 0.137   | 0.1355        |
| $\sigma$  | 0.6     | 1.0     | 0.66667       |
| $c_{b2}$  | 0.60983 | 0.6875  | 0.622         |
| $K$       | 0.38    | 0.46    | 0.41          |
| $c_{w2}$  | 0.055   | 0.3525  | 0.3           |
| $c_{w3}$  | 1.75    | 2.5     | 2.0           |
| $c_{t1}$  | 6.9     | 7.3     | 7.1           |
| $c_{t3}$  | 1.0     | 2.0     | 1.2           |
| $c_{t4}$  | 0.3     | 0.7     | 0.5           |
Table 2. The statistical information of the lift and drag coefficients obtained by different models

|                | Lift coefficient | Drag coefficient |
|----------------|------------------|------------------|
|                | mean             | MSE              | mean             | MSE              |
| low-fidelity   | 0.67879          | 4.5254E-3        | 1.4968E-2        | 2.9517E-4        |
| high-fidelity  | 0.69601          | 3.9577E-3        | 1.2973E-2        | 2.6323E-4        |
| modified model | 0.69596          | 4.0701E-3        | 1.2976E-2        | 2.6411E-4        |

Table 3. The number of deterministic CFD calculations performed by the two high-fidelity models

|                | calculation times (original grid) | calculation times (coarser grid) |
|----------------|----------------------------------|----------------------------------|
| modified model | 20                               | 110                              |
| high-fidelity  | 110                              | 0                                |

In order to further verify the accuracy of the modified model, Figure 3 and Figure 4 respectively show the comparison of the lift, drag coefficient and CFD calculation value predicted by the two models at the sample points used when directly building the model. The closer the discrete points are to the green straight line, the higher the accuracy of the model. It can be clearly seen from the figure that the accuracy of the two models is relatively high.

Table 4 shows the $R^2$ coefficients of determination of the two models, which are respectively defined on two sets of sample points for establishing the model directly and establishing the model through a modified method. The $R^2$ coefficient of determination of the two models is very close to 1, indicating that both models have high accuracy. On the sample points of the directly established model, the directly established high-fidelity model has higher accuracy, and on the sample point of the model established by the modified method, the modified model has higher accuracy. In general, the test accuracy of the modified model on the two sets of sample points is in line with the requirements, which can be said to be equivalent to the accuracy of the directly established high-fidelity model, which also proves that this method can save computing resources.

Table 4. $R^2$ determination coefficient of high-fidelity and modified models

|                | directly established model | modified method |
|----------------|---------------------------|-----------------|
|                | lift coefficient          | drag coefficient| lift coefficient| drag coefficient|
| modified model | 0.9912                    | 0.9983          | 0.9994          | 1.0000          |
| high-fidelity  | 0.9997                    | 0.9998          | 0.9986          | 0.9999          |

Figure 3. Comparison of lift coefficients and calculated values predicted by the two models (Left: High-fidelity model, Right: modified model)
Figure 4. Comparison of drag coefficients and calculated values predicted by the two models (Left: High-fidelity model, Right: modified model)

In the case of multiple input variables, the uncertainty of each variable contributes differently to the uncertainty of the system output. The Sobol indices provide tools to quantify the contribution of each input variable to the system output variance. The Sobol indices are obtained by Sobol decomposition. It is a global sensitivity analysis method based on variance decomposition. Its advantage is that it considers the contribution of random input parameters to the output response in the entire range of values and the interaction of random input parameters. Table 5 shows the Sobol indices of each of the 9 parameters in the prediction of the lift and drag coefficients obtained by the three models of low-fidelity model, high-fidelity model and modified model. In this example, there is almost no difference in the Sobol indices obtained by the three models. It should be noted that the Sobol indices is dimensionless by the output variance. In the previous article, it was pointed out that the variance between the low-fidelity model and the high-fidelity model is obviously different, so the degree of freedom of the low-fidelity PC expansion is also significantly different from the degree of freedom of the high-fidelity model.

Through detailed comparison of the output statistics, including the mean, variance, and global sensitivity of each input parameter, it can be found that the modified model is equivalent to the directly established high-fidelity model, but the amount of calculation is reduced about 60%. This is a method that can handle the quantification of multivariate uncertainties in engineering.

Table 5. Sobol Indices of Closure Coefficients for lift and drag coefficients

| model para | Lift coefficient | Drag coefficient |
|------------|------------------|------------------|
|            | low-fidelity     | modified         | low-fidelity     | modified         |
|            | high-fidelity    | model            | high-fidelity    | model            |
|            | Sobol indices    |                  | Sobol indices    |                  |
| κ          | 7.005×10-1       | 7.006×10-1       | 5.799×10-1       | 5.799×10-1       |
|            | 7.004×10-1       | 7.006×10-1       | 5.799×10-1       | 5.799×10-1       |
|            | 5.799×10-1       | 5.799×10-1       | 5.799×10-1       | 5.799×10-1       |
| c1         | 1.177×10-1       | 1.177×10-1       | 2.554×10-1       | 2.554×10-1       |
|            | 1.177×10-1       | 1.177×10-1       | 2.554×10-1       | 2.554×10-1       |
|            | 2.554×10-1       | 2.554×10-1       | 2.554×10-1       | 2.554×10-1       |
| σ          | 8.727×10-2       | 8.727×10-2       | 1.071×10-1       | 1.071×10-1       |
|            | 8.727×10-2       | 8.727×10-2       | 1.071×10-1       | 1.071×10-1       |
|            | 1.071×10-1       | 1.071×10-1       | 1.071×10-1       | 1.071×10-1       |
| σ1         | 7.706×10-2       | 7.706×10-2       | 3.013×10-2       | 3.013×10-2       |
|            | 7.706×10-2       | 7.706×10-2       | 3.013×10-2       | 3.013×10-2       |
|            | 3.013×10-2       | 3.013×10-2       | 3.013×10-2       | 3.013×10-2       |
| σ2         | 1.649×10-2       | 1.649×10-2       | 2.840×10-2       | 2.840×10-2       |
|            | 1.649×10-2       | 1.649×10-2       | 2.840×10-2       | 2.840×10-2       |
|            | 2.840×10-2       | 2.840×10-2       | 2.840×10-2       | 2.840×10-2       |
| σ3         | 5.567×10-3       | 5.567×10-3       | 6.884×10-4       | 6.884×10-4       |
|            | 5.567×10-3       | 5.567×10-3       | 6.884×10-4       | 6.884×10-4       |
|            | 6.884×10-4       | 6.884×10-4       | 6.884×10-4       | 6.884×10-4       |
| σ4         | 3.101×10-4       | 3.101×10-4       | 2.735×10-4       | 2.735×10-4       |
|            | 3.101×10-4       | 3.101×10-4       | 2.735×10-4       | 2.735×10-4       |
|            | 2.735×10-4       | 2.735×10-4       | 2.735×10-4       | 2.735×10-4       |
| σ5         | 2.284×10-4       | 2.284×10-4       | 1.881×10-4       | 1.881×10-4       |
|            | 2.284×10-4       | 2.284×10-4       | 1.881×10-4       | 1.881×10-4       |
|            | 1.881×10-4       | 1.881×10-4       | 1.881×10-4       | 1.881×10-4       |
| σ6         | 1.134×10-4       | 1.134×10-4       | 4.689×10-5       | 4.689×10-5       |
|            | 1.134×10-4       | 1.134×10-4       | 4.689×10-5       | 4.689×10-5       |
|            | 4.689×10-5       | 4.689×10-5       | 4.689×10-5       | 4.689×10-5       |

4. Conclusion
In this paper, a high-fidelity polynomial chaos method suitable for CFD uncertainty quantification is established and verified in the RAE2822 airfoil calculation. It is proved that this method can reduce the
calculation amount by 60% while ensuring the calculation accuracy. Greatly improve the efficiency of uncertainty quantification.

Acknowledgments
This work was financially supported by National Numerical Wind Tunnel Project.

References
[1] Wiener N, Am J Math. 60(1938) 897.
[2] Xiu D, Karniadakis G, SIAM J Sci Comput. 24(2002) 619.
[3] Sobol’, I, Math. Modeling & Comp. Exp. 1(1993) 407.
[4] Cook P. H, McDonald M. A, etc. AGARD Re-port AR 138, 1979.
[5] J T Chen, Y B Zhang, H Zhao, et al, Journal of Aircraft, 55 (2018) 1469.