DOI: 10.13009/EUCASS2017-108

Aerodynamic Data Fusion with a Multi-fidelity Surrogate Modeling Method

*HE Lei *,**, ZHOU Yu *,** , QIAN Wei-qi *,** and WANG Qing*,***
** China Aerodynamics Research and Development Center*
Mianyang Sichuan 621000, China
*** State Key Laboratory of Aerodynamics*
Mianyang Sichuan 621000, China

Abstract
The aerodynamic data fusion method with multi-fidelity surrogate proposed in this paper provides an effective way to generate high quality aerodynamic data. RBF and Kriging surrogate modelling methods are discussed. RBF is used to build the low-fidelity model due to its high efficiency in practice and Kriging is used to build increment model due to its high accuracy. The difficulties of matrix operation that may occur during the process of constructing the surrogate model with large scale low-fidelity data set are also discussed and the solution is proposed later. Two study cases are taken as example to validate the data fusion method, one is about aerodynamic force, and another one is about aerodynamic heat flow. The fusion results are both satisfactory. It can be concluded that the method proposed in this paper can successfully deal with aerodynamic data fusion problem.
Keyword： Aerodynamic Data Fusion; Multi-fidelity; RBF; Kriging

1. Introduction

In general, aerodynamic data is generated via three sorts of aerodynamic testing: flight testing, wind-tunnel testing, and computational simulation[1]. Flight testing can obtain the most accurate and reliable aerodynamic data and is often used as a final assessment, however, the flight testing is so expensive and the test cycle so long. So, during the engineering development phase, aerodynamic development and analysis rely on wind-tunnel testing. Wind-tunnel testing is an important means to simulate the performance of aircraft, including the prediction of aerodynamic force/heat inside the flight envelope, the establishment of aerodynamic database, confirming the reliability of numerical simulation results[2]. Although the test data accuracy is relatively accurate, wind-tunnel testing is not cheap to perform in terms of cost, time and resources. Computational Fluid Dynamics (CFD) is also widely applied in the aerodynamic engineering field to simulate the performance of systems, saving the cost of expensive wind tunnel experiments. Especially in recent years, benefit from the development of computer technology, the ability of the CFD method to simulate the complex physical problem have rapid and constantly improved, however, such computational experiments still consume time to produce a large number of accurate aerodynamic data and often show discrepancies in results compared with experiments, particularly with low-fidelity computational models. Therefore, CFD method is commonly conducted at an initial stage of aerodynamic development and often be examined by wind-tunnel testing. It is often necessary to combine experimental and computational aerodynamic data sets to successfully produce high-quality aerodynamic data[3,4].
The aerodynamic characteristics are closely related to its configuration and flow field. Compared with the conventional sensor measurement data, aerodynamic data has its own complexity and particularity. Modeling-based aerodynamic data fusion method usually needs to construct the mathematical model which can describe the aerodynamic characteristics and the data fusion operation bases on. Data modeling methods are often divided into two groups: conventional aerodynamic modeling method with explicit physical meaning and surrogate-based aerodynamic modeling method[5]. The aerodynamic data fusion steps based on conventional model with explicit physical meaning are: the analysis of the aircraft aerodynamic characteristics, the construction of aerodynamic model, the parameter identification of model using data from different sources. The advantages of this method are that the model can reflect the aerodynamic characteristics in the level of physical laws and the model has the anti-interference ability on the aerodynamic noise data. The deficiencies are that the modeler must have a deep understanding on the physical theory of aerodynamic and the modeling period is always too longer. Surrogate-based aerodynamic data fusion method needs to construct a surrogate model. Although the surrogate model is more

DOI: 10.13009/EUCASS2017-108

sensitive to noise data, which cause that its anti-interference ability is inferior to the former. However, its modeling efficiency is higher and modeling time is shorter. What's more, with the development of computer technology and wind tunnel test technology, the quality of the aerodynamic data continuously increase, the interference amplitude noise data continuously decrease.

In this study, we use a multi-fidelity surrogate modeling method to fuse the low-fidelity sample data of the computations with a few high-fidelity data of the wind tunnel experiments. The target of method is to generate dataset that is more accurate than the low fidelity data and more quantity than high fidelity data. The idea of the method relies on the assumption that low fidelity data is used to predict global trends while high fidelity data is used to provide absolute values and correct the global trends. Surrogate modelling is widely used to describe the performance of system, due to its simplicity and efficiency[6,7]. Surrogate model is an approximation of the relationship between the input variable and their response of certain system, and it often be used in the initial development phases to reduce the resource required for design and optimization[8,9]. Surrogate models are generally constructed by two steps: (a) choose a predefined function according the design points in the region of interests, (b) determine coefficients of the predefined function by the true values of system at some design points.

2. Surrogate Models

Surrogate models is often applied in the domain of engineering design and optimization of aerospace, which can statistically approximate the relationship between a set of design variables and their response, resulting in reducing the resource required for design, search and optimization. Consider an input-response model which commonly represented mathematically as[10]:

$$\boldsymbol{E[Y] = \eta = g(X)} \tag{1}$$

Be sure that the symbols in your equation are defined before or immediately following the equation or add a nomenclature. where $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{p})^\mathrm{T}$ is a vector of p input variables or design variables over the experimental region D and η is the output for a given input vector x from the true response function g(·). For most engineering applications, g(·) is an unknown function, Therefore, it is desirable to construct an approximate function of g(·) that is explicit, less computationally expensive, and accurate enough for the entire design space. The unknown g(x) is approximated by a function $\hat{g}(X; \beta)$, where β is a vector of unknown coefficients, and the primary task then becomes one of estimating the coefficients.

Surrogate Models include polynomial response surface (PRS), spatial correlation models (or "kriging"), Multivariate Adaptive Regression Splines (MARS), regression trees and boosting, Artificial Neural Networks (ANNs), Radial Basis Functions (RBFs), and least interpolating polynomials[11,12]. For all approaches, the general task at hand is to estimate the relationship between a response variable and several inputs. With the known values of g(x) corresponding to a given set of design variables, the approximate function can be constructed using the RSM, RBF, Kriging, or other methods. This study focused on constructing surrogate models using the Kriging and RBF.

2.1 Kriging model

The Kriging model assumes some form of spatial correlation between points in the multi-dimensional space, and uses this correlation to predict response values between observed points[13,14]. The general form can be written as:

$$\boldsymbol{Y(X) = \sum_{m=1}^{M} \beta_m\, b_m(X) + Z(X)} \tag{2}$$

where $\sum_{m=1}^{M} \beta_m\, b_m(X)$ is the linear regression model of M known functions $b_m(X)$ with unknown coefficients β_m. The stochastic component Z(x) is a random process, commonly assumed to be Gaussian, with mean zero and covariance $Cov[Z(X), Z(X')] = \sigma^2 R(X, X')$, where σ^2 is the process variance and $R(X, X')$ is the correlation. The most common form of correlation function is:

$$\boldsymbol{R(X, X') = \prod_{j=1}^{p} e^{-\theta_j \left|x_j - x_j'\right|^{p_j}}} \tag{3}$$

where $\theta_j \geq 0$, $1 \leq p_j \leq 2$. The "Gaussian" correlation function ($p_j = 2$) results in an infinitely differentiable approximation appropriate for smooth underlying functions[15]. It is the hyperparameters, θ_j, that are optimized to train the model. This is performed by maximizing the following logarithmic likelihood.

DOI: 10.13009/EUCASS2017-108

$$\max_{\theta_j>0} \varphi(\theta_j) = {}^{1}\!/_{2}\,[n_s \ln \hat{\sigma}^2 + \ln|\boldsymbol{R}|] \quad \textbf{(4)}$$

In practice, a constant regression model is commonly used as shown in Eq.(5), since the inclusion of a more complex linear model does not necessarily yield a better prediction.

$$\boldsymbol{Y}(\boldsymbol{X}) = \boldsymbol{\beta} + \boldsymbol{Z}(\boldsymbol{X}) \quad \textbf{(5)}$$

For a given set of sample points $\{(X^1, Y^1), (X^2, Y^2), \cdots, (X^{n_s}, Y^{n_s})\}$, the approximation $\widehat{\boldsymbol{Y}}$ of the input vector $\boldsymbol{X}$ is:

$$\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\beta}} + \boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{X})\boldsymbol{R}^{-1}(\boldsymbol{Y}_{\boldsymbol{s}} - \boldsymbol{F} \cdot \widehat{\boldsymbol{\beta}}) \quad \textbf{(6)}$$

where $\boldsymbol{Y}_{\boldsymbol{s}} = [Y^1, Y^2, \cdots, Y^{n_s}]$, F $= (1,1,\cdots,1)^T$ and

$$\boldsymbol{r}(\boldsymbol{X}) = [\boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{1}), \boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{2}), \cdots, \boldsymbol{R}(\boldsymbol{X}, \boldsymbol{X}^{n_s})] \quad \textbf{(7)}$$

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}(\boldsymbol{X}^1, \boldsymbol{X}^1) & \cdots & \boldsymbol{R}(\boldsymbol{X}^1, \boldsymbol{X}^{n_s}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}(\boldsymbol{X}^{n_s}, \boldsymbol{X}^1) & \cdots & \boldsymbol{R}(\boldsymbol{X}^{n_s}, \boldsymbol{X}^{n_s}) \end{bmatrix} \quad \textbf{(8)}$$

β is estimated using the method of least squares,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}^{\boldsymbol{T}}\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^{\boldsymbol{T}}\boldsymbol{R}^{-1}\boldsymbol{Y}_{\boldsymbol{s}} \quad \textbf{(9)}$$

And the estimated variance value of $\hat{\sigma}^2$ can be obtained as,

$$\hat{\sigma}^2 = [(\boldsymbol{Y}_{\boldsymbol{s}} - \boldsymbol{F}\hat{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{Y}_{\boldsymbol{s}} - \boldsymbol{F}\hat{\beta})]/n_s \quad \textbf{(10)}$$

2.2 RBF model

The general form of the RBF approximation can be written as:

$$\hat{\boldsymbol{y}}(\boldsymbol{X}) = \sum_{i=1}^{n} \boldsymbol{\beta}_i \boldsymbol{b}(\|\boldsymbol{X} - \boldsymbol{X}_i\|) \quad \textbf{(11)}$$

where n is the number of the response observations, $\|X - X_i\|$ is the Euclidean norm between the design variable X and the i^{th} observation point, and β_i is the corresponding weight coefficient of the i^{th} basis function $b(\|X - X_i\|)$. A matrix equation denoted as below is used to solve for the coefficients $\boldsymbol{\beta}$:

$$\boldsymbol{A\beta} = \boldsymbol{Y} \quad \textbf{(12)}$$

Where $A_{ij} = b(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|)$ and $\boldsymbol{Y}$ is the vector of response observations. Several choices for the basis function b(·) are summarized in table 1[16,17]:

Table 1: Basis functions

Name	Basis function
Linear	$b(r) = r$
Cubic	$b(r) = r^3$
Gaussian	$b(r) = exp(-cr^2),\ 0<c\leqslant 1$
Multi-quadratic	$b(r) = (r^2 + c^2)^{1/2},\ 0<c\leqslant 1$
Inverse multi-quadric	$b(r) = (r^2 + c^2)^{-1/2},\ 0<c\leqslant 1$
Thin plate spline	$b(r) = r^2 \ln (cr),\ 0<c\leqslant 1$

As to the basis function, Gaussian basis is commonly used. As described above, the method is an interpolating approximation. RBF approximations have produced good fits to arbitrary contours of both deterministic and stochastic responses[18].

DOI: 10.13009/EUCASS2017-108

2.3 Comparisons between kriging and RBF model

Both Kriging and RBF method are well global interpolation technique, Ai Yi Si made comparison between the two methods[19], and the results are displayed in table 2 and table 3. Root mean square error (RMSE) is used to describe the accuracy of the surrogate model, and the construct time is used to describe how difficulty to construct the surrogate model. Six function has been used to test，including Six-hump Camel-Back Function (SCBF), Banana Function (BF), Schaffer’s Function (SF), Generalized polynomial function (GF), Goldstein and Price Function (GP) , and all test points are sampled at the same experimental region [-2,-2; 2,2].

Table 2: Comparison of interpolation accuracy between Kriging and RBF

Sample data	20		200		2000	
RMSE	Kriging	RBF	Kriging	RBF	Kriging	RBF
SCBF	3.8779	5.725	0.0045	0.3565	2.8067e-005	0.0092
BF	104.9760	112.7978	0.0010	9.1819	5.6382e-005	0.4787
SF	0.1154	0.084	1.0026e-006	0.0071	1.2390e-008	2.0347e-04
GF	41.4654	22.8205	0.0104	2.5423	6.0568e-005	0.1024
GP	2.0080e+05	7.61e+04	10.4108	6.808e+03	0.5696	194

Table 3: Comparison of interpolation time between Kriging and RBF

Sample data	20		200		2000	
time	Kriging	RBF	Kriging	RBF	Kriging	RBF
SCBF	0.016148	0.0033	0.411711	0.1007	59.797163	79.2178
BF	0.018040	0.0014	0.316916	0.1099	54.598063	87.112
SF	0.015633	0.0014	0.413025	0.0972	55.994902	78.0021
GF	0.021908	0.0019	0.462016	0.1004	60.292655	79.6166
GP	0.016101	0.0034	0.406938	0.101	62.856208	80.2997

We can conclude from the test results: the addition of data points in a Kriging model would always increase the accuracy of the interpolation model, but the time to construct the model also constantly increases. The RBF interpolating approximation’s complexity of computation is lower than Kriging with medium or small data sets, although its accuracy is lower than Kriging at the sample points with big data sets. So RBF surrogate model can reduce computation expense compared with Kriging with small data sets, and guarantee the accuracy at the same time. The RBF surrogate model also has some shortages: the selection of basis-functions and related parameters according to the system characteristics and experience is always not easy; surrogate model is constructed slowly with big sampling dataset; the accuracy of surrogate model depends on the selection of sample points.

3. Aerodynamic Data Fusion with a Multi-fidelity Surrogate Modeling Method

The idea of data fusion is to merge low fidelity data with a few, high quality solutions so the resulting dataset is more accurate than either the low or high fidelity data separately. Essentially, it relies on the assumption that low fidelity data is used to predict trends while high fidelity data is used to provide absolute values. To implement this idea, two function are commonly used, namely, scaling function and increment function. The scaling function $\sigma(x)$ define the ratio between a high fidelity (f_{hf}) and a low fidelity (f_{lf}) solution[20].

$$\boldsymbol{\sigma(x_s) = f_{hf}(x_s)/f_{lf}(x_s)} \tag{13}$$

where $\boldsymbol{x_s}$ is the input variables at the observation points. The values of this scaling function $\sigma(x)$ are interpolated throughout the whole design space. Then the function was approximated using the expression:

$$\boldsymbol{f(x) = \sigma(x) f_{lf}(x)} \tag{14}$$

However there are some potential problems associated with scaling function approach when it combines low and high fidelity data. If the value of the low fidelity data is exactly zero, the approach would not work. And if the low fidelity data is close to zero, $\sigma(x)$ may be quite large and amplify any approximation errors. To avoid these possible

DOI: 10.13009/EUCASS2017-108

problems, the increment function β(x) is proposed to instead of computing the ratio between the high and low quality data[21].

$$\boldsymbol{\beta}(\boldsymbol{x_s}) = \boldsymbol{f_{hf}}(\boldsymbol{x_s}) - \boldsymbol{f_{lf}}(\boldsymbol{x_s}) \tag{15}$$

Similarly, the function is approximated using the equation:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{\beta}(\boldsymbol{x}) + \boldsymbol{f_{lf}}(\boldsymbol{x}) \tag{16}$$

This increment function is more reliable than a scaling ratio since the subtraction of small values does not result in any amplification errors.
To build the increment model, we can use the Kriging method to construct the surrogate model. Kriging surrogate model goes through the data points exactly, and can approximate the true model with fewer data[22]. It can be guaranteed that fusion results are equal to the high fidelity data completely and the quality of model with small sample data sets. So the Kriging method is appropriate to build the increment model.
To build the low-fidelity model, some studies also use the Kriging surrogate modeling method. Despite it has the advantage of accuracy and robustness with small data sets and the addition of data points in a Kriging model would always increase the accuracy of the interpolation model, the time-consuming will increase significantly. Particularly, it is difficult for ordinary computers to support the matrix operations required by the Kriging model as the amount of data points increase to a large scale. There are a lot of matrix operations that exist in the process to build a surrogate model with Kriging method, especially the inverse operation of the covariance matrix R. If the amount of data points used to build Kriging surrogate model is n, the size of the covariance matrix R is $n \times n$. In engineering practice, the amount of low-fidelity data points often reach tens of thousands or even more, so the covariance matrix involved in model-built process would be even large and the matrix operation would be more difficult. Therefore, it is not appropriate to build a surrogate model with the whole data set directly when the amount of low fidelity data is large. To solve this problem, we can choose a small part of low fidelity data points to build this model, all those points selected should be distributed around the location of the point that will be predicted. The method how to select the low fidelity data points will be discussed in section 4.2.As a result, the low-fidelity model will be built repeatedly. Compared to the Kriging method, it is less time-consuming, more simple and easier to construct a surrogate model with RBF method. In addition, it is easier to adjust the smoothness of the RBF surrogate model[23], and the approximate accuracy and the computational complexity is suitable. So the RBF method can be used to build the low-fidelity model.
According the analysis above, the process of data fusion based on surrogate model is:

(1) Select one data point [$\boldsymbol{x}_h$ y_h] in order from the high-fidelity data set;

(2) If there is low-fidelity data value y_l that exist at the high-fidelity data point position $\boldsymbol{x}_h$, then it can be used directly. If not, to obtain the value by choosing a small part of low fidelity data points to build low-fidelity model with RBF method.

(3)To calculate the increment β between low-fidelity data value and high-fidelity data value, where $\beta = y_h - y_l$.

(4) Repeat steps (1) - (3) until the increment values at all the high-fidelity data positions are calculated.

(5) Build the increment model based on the Kriging method.

The result of fusion is the summation of calculate values of the low-fidelity model and values of the increment model. For any point in the test area, the fusion method can be used to predict its value. Practically, when the low-fidelity data set is small, the low-fidelity model can be built with whole data set directly.

4. Example demonstrations

In this section, two cases are used to verify the proposed fusion method. Root mean square error (RMSE) is used to describe the accuracy of the results, defined below:

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^2} \tag{17}$$

where n represents the total number of validation points, and y_i is the real response at the validation point while $\hat{y}_i$ is the predicted value.

DOI: 10.13009/EUCASS2017-108

4.1 Aerodynamic force data fusion case

Taking the aerodynamic force data fusion of an aircraft as an example. The fusion data include aerodynamic force data obtained from two different wind tunnel test conditions. One data set is called basis data and another is called object data, and a part of data points selected from object data are called sample data.The test variables of aerodynamic force data are the attack angle and sideslip angle (α, β), the responses are six aerodynamic force components $(C_A, C_Y, C_N, C_l, C_m, C_n)$. Normal force coefficient (C_N) is used here to test the fusion method above. As illustrated in Figure 1, the green line is object data of normal force coefficient and black line is basic data of normal force coefficient. A part of data points are selected from the object data set as the sample data. As shown in Figure 2, the red points are sample data. There are 75 aerodynamic force data points with different experimental states of attack angle and sideslip angle for both basic data and object data. Where,

$$\begin{cases} \boldsymbol{\alpha} = [-4^o, -2^o, 0^o,\ 1^o,\ 2^o\ , 3^o, 4^o, 5^o, 6^o, 8^o, 9^o, 10^o, 11^o, 13^o, 15^o] \\ \boldsymbol{\beta} = [0^o, 4^o, 8^o, 12^o, 16^o] \end{cases} \tag{18}$$

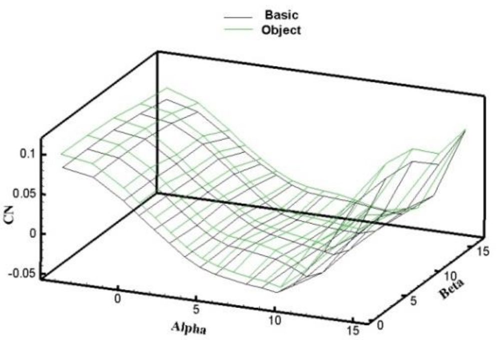
The samples selected from object data points contain 24 aerodynamic force data points with different experimental states of attack angle and sideslip angle, where

$$\begin{cases} \boldsymbol{\alpha} = [-4^o, 0^o,\ 3^o\ , 5^o, 7^o, 9^o, 12^o, 15^o] \\ \boldsymbol{\beta} = [0^o, 8^o, 16^o] \end{cases} \tag{19}$$

The basic data can be used as low-fidelity data and the sample data can be used as high-fidelity data. According to the method discussed above, we use RBF surrogate to model the basic data and Kriging surrogate to model sample data, and start the data fusion process. Data fusion results are shown in Figure 3. The fusion results (red line), object data, sample data are all include in figure 3. Obviously, fusion data and sample data of normal force coefficients (C_N) coincide exactly, and in other positions of the test region, the fusion data reflect the trend of object data and can coincide in most test region. Besides, the root mean square error of the fusion results is:

$$\boldsymbol{RMSE} = \mathbf{0.0025675} \tag{20}$$

It indicates the error is small and the accuracy of the results is satisfactory. So the results reveal that the fusion model is effective.



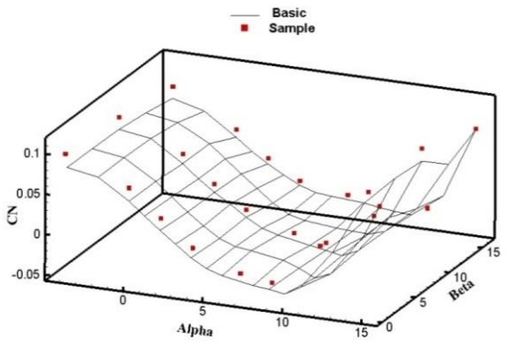


Fig.1 Basic data and object data of $\mathrm{C_N}$ Fig.2 Basic data and sample data of $\mathrm{C_N}$

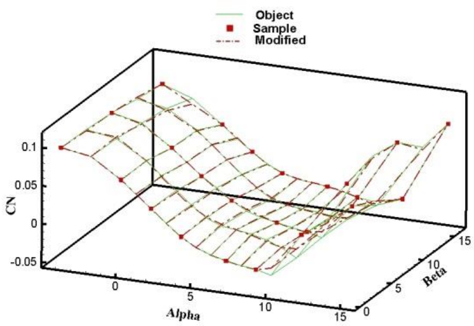


Fig.3 Comparison of $\mathrm{C_N}$ between object data and fusion results

DOI: 10.13009/EUCASS2017-108

4.2 Aerodynamic heat data fusion case

Taking the aerodynamic heat flow data fusion of an aircraft as an example. The fusion data include the continuous heating environment data from CFD numerical simulation and the heat flow values obtained from of the wind tunnel test points. There are some requirements here: fusion results are coincident with measured results in the locations of wind tunnel test points, the CFD numerical simulation values are appropriate corrected, and the data fusion results can reflect the characteristic of wind tunnel test data on the whole test region. All the positions where we gain the aerodynamic heat data are illustrated in fig.4. 32922 blue points were computed using CFD software to gain the aerodynamic heat data as the low fidelity data, and the result is shown in fig5. 118 red points were experimented through wind tunnel to gain the aerodynamic heat data as the high fidelity data, and the result is shown in fig.6.

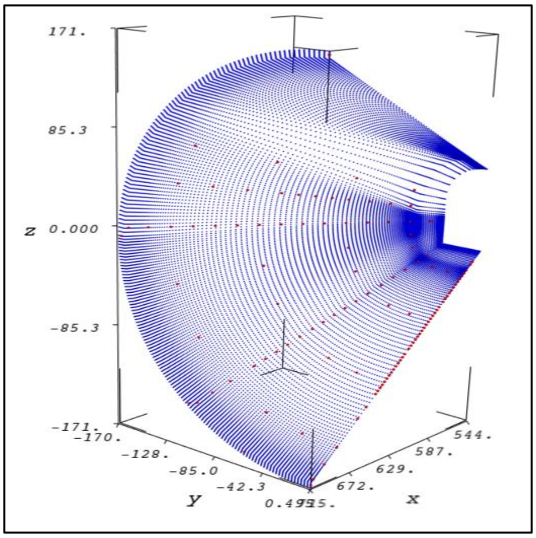
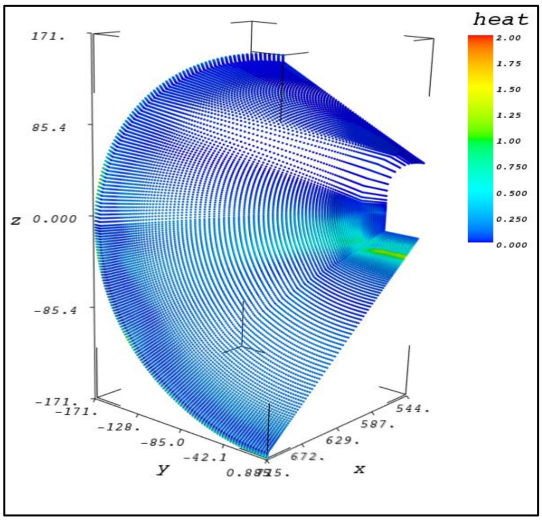


Figure 4: Position distribution of CFD and Wind Tunnel



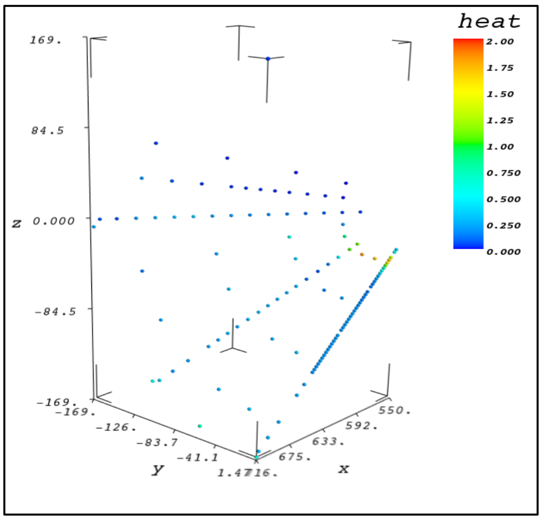


Figure 5: CFD numerical calculation data

Figure 6: Wind tunnel test data

In the process of data fusion, increment model ($\beta(x)$) of heat data from wind tunnel test and CFD numerical calculation should be built, but wind tunnel data and CFD numerical calculation data are usually not available at the same locations, so it is necessary to interpolate the CFD numerical calculation data at the locations of all the wind tunnel test data that used to modeling. However, there are more than 30000 heat flow data of CFD numerical calculation, and just as discussed before, it is difficult to establish a RBF surrogate mode for the whole data. To solve this problem, we can choose a small part of CFD numerical calculation data around the location of the wind tunnel test data each time to build the low-fidelity model with RBF method. The low-fidelity heat flow value at the location of the wind tunnel test data can obtain based the low-fidelity model.

DOI: 10.13009/EUCASS2017-108

The relative position of the CFD numerical calculation data points can be described by array [*I*, *J*], since the data are copulated at structural gird points. Then, the CFD data set we choose around the wind tunnel test data can be easily described. The method is shown in Figure 7, where the blue dot represents the CFD grid point position and the red dot represents the wind tunnel test point position. For example, In Figure 7a, the wind tunnel test point is located between CFD data points [4, 4] [4, 5] [5, 4] and [5, 5]. In order to improve the modeling accuracy, we determine the modeling data by 4 steps both in *I* and *J* axis, namely, *I* = [3, 4, 5, 6], *J* = [3, 4, 5, 6], so 16 CFD data are selected in all to modeling. In Figure 7b, the wind tunnel test point is located in the edge of test area. Also, we determine the CFD data point by 4 steps, then the condition that *I* = 9 appears. The state that I = 9 should be ignored, because there are no points in that position. So 12 CFD data are selected in all to modeling, where *I* = [6, 7, 8], *J* = [3, 4, 5, 6].

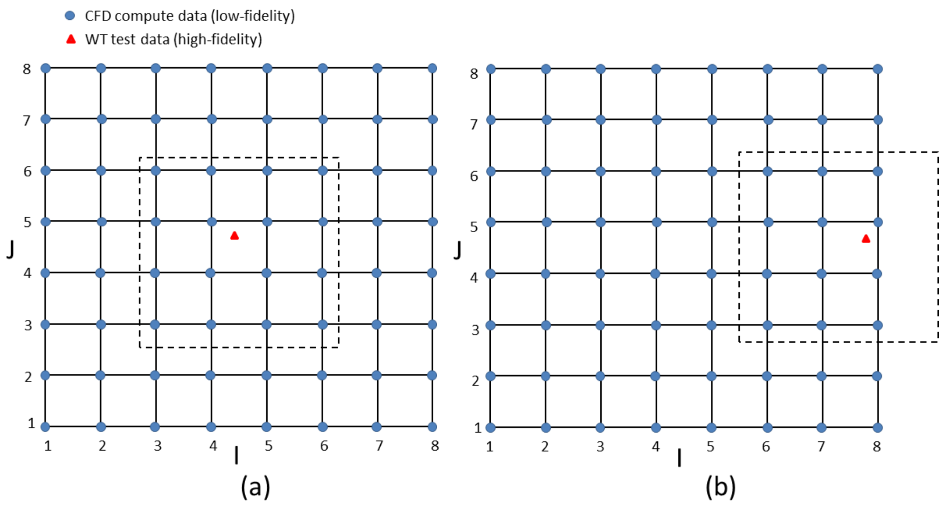
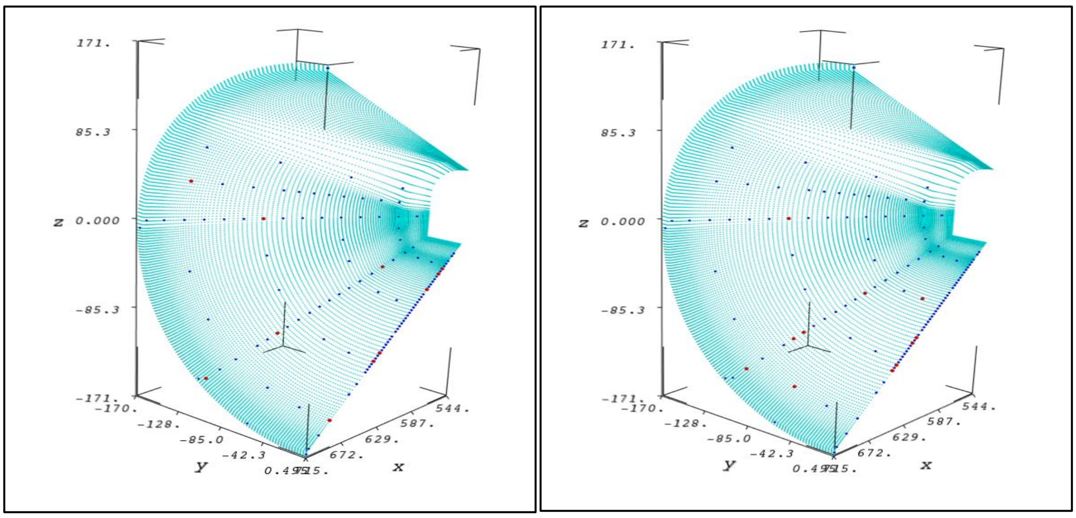


Figure 7: Strategy for choosing low-fidelity data points

In order to validate the fusion method proposed in this paper, about 10% of the wind tunnel test data, totally 11 data points, were randomly selected as validation data, the remaining 90% were used to construct the fusion model with all CFD numerical calculation data. The red dots in Figure 8a and figure 8b are the 10% randomly selected points from the wind tunnel test data. Figure 9a and figure 9b are the comparison between data fusion results and validation data, where horizontal axis the true value of the wind tunnel test data for the validation and vertical axis is the predicted values obtained from the fusion model at the same locations of validation data.



(a) (b)

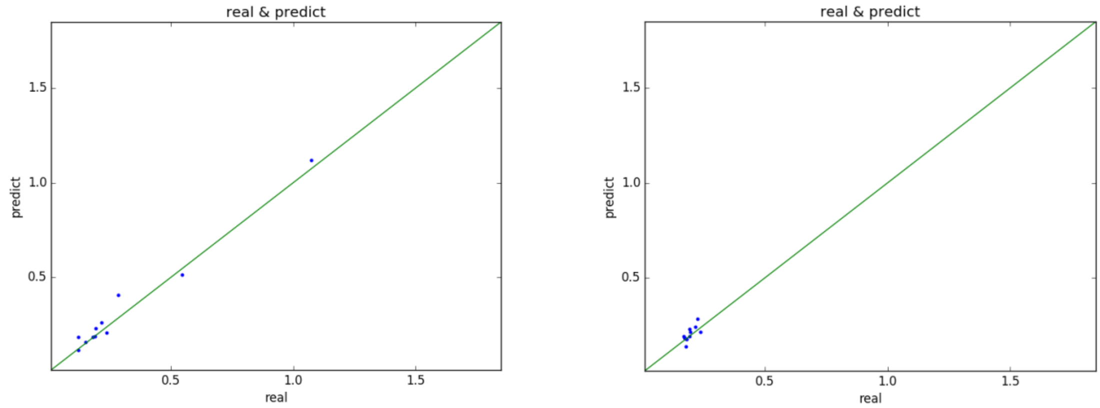
Figure 8: Position distribution of validation data

The root mean square error of the two cases:

DOI: 10.13009/EUCASS2017-108

$$\begin{cases} \boldsymbol{RMSE}_{\ (a)} & = \ \mathbf{0.04782} \\ \boldsymbol{RMSE}_{\ (b)} & = \ \mathbf{0.02762} \end{cases} \tag{21}$$

It indicates the error is small and the values predicted are very close to the real values. So the conclusion can be drawn that the fusion model is effective and the accuracy of the results is satisfactory



(a) (b)

Figure 9: The comparison of values between fusion results and validation data

5. Conclude

It has great potential and prospects of data fusion technology in the domain of aerodynamic. The aerodynamic data fusion method with multi-fidelity surrogate proposed in this paper provides an effective way to generate high quality aerodynamic data. It can also reduce the number of wind tunnel test (or flight test) and save cost. In this paper, two surrogate modeling methods, namely, RBF and Kriging, are discussed, and the advantages and shortage of them are analyzed. RBF is used to build the low-fidelity model due to its high efficiency in practice and Kriging is used to build increment model due to its high accuracy. Two study cases are taken as example to validate the data fusion method, one is about aerodynamic force, and another one is about aerodynamic heat flow. In the aerodynamic heat flow case, solution is proposed to solve the difficulties of matrix operation with large scale low-fidelity data set. From the value of RMSE and the comparison between fusion results and validation data, it can be concluded that the results are satisfactory.

References

[1] Chen Jianqiang, Zhang Y R, Zhang Y F, Chen LZ. 2014. Review of correlation analysis of aerodynamic data between flight and ground prediction for hypersonic vehicle. ACTA Aerodynamic Sinica. 32(5) :587-559

[2] Hall, R. M., Biedron, R. T., Ball, D. N., Bogue, D. R., Chung, J., Green, B. E., Grismer, M. J., Brooks, G. P., and Chambers, J. R., Computational Methods for Stability and Control (COMSAC) The Time Has Come, AIAA Atmospheric Flight Mechanics Conference and Exhibit, San Francisco, California, 15 - 18 August 2005 2005.

[3] Y. Kuya, K. Takeda, and X. Zhang. 2009.Optimal Surrogate Modeling Approaches for Combining Experimental and Computational Fluid Dynamics Data Sets. AIAA 2009-2216

[4] Y. Kuya, K. Takeda, X. Zhang, and Alexander I. J. Forrester. 2011. Multi-fidelity Surrogate Modeling of Experimental and Computational Aerodynamic Data Sets. AIAA JOURNAL

[5] He Kaifeng, Qian Weiqi, Wang Qing, Kong Yinan, Wang Wenzheng. 2014. Application of data fusion technique in aerodynamics studies. ACTA Aerodynamic Sinica. 32(6):777-782.

[6] J. Laurenceau and P. Sagaut, Building efficient response surfaces of aerodynamic functions with kriging and cokriging, AIAA J.46(2), 498–507 (2008).

[7] A. I. J. Forrester, A. Sobester, and A. J. Keane, Engineering Design via Surrogate Modelling – A Practical Guide (Wiley, New York, 2008).

[8] T. J. Santner, B. J. Williams, and W. Notz, The Design and Analysis of Computer Experiments, in: Springer Series in Statistics, 1st edition (Springer, Berlin, Heidelberg, New York, 2003).

[9] Chanyoung Park, Raphael T. Haftka, Nam H. Kim. 2016. Remarks on multi-fidelity surrogates. Structural and Multidisciplinary Optimization

DOI: 10.13009/EUCASS2017-108

[10] Victoria C.P. Chen , Kwok-Leung Tsui , Russell R. Barton & Martin Meckesheimer (2006). A review on design, modeling and applications of computer experiments, IIE Transactions, 38:4, 273-291, DOI: 10.1080/07408170500232495

[11] Russell R. Barton. Simulation optimization using metamodels, Proceedings of the 2009 Winter Simulation Conference,1(4):230-238

[12] Younis, A., and Z. Dong. 2010. Trends, Features, and Tests of Common and Recently Introduced Global Optimization Methods. Engineering Optimization 42 (8): 691–718. doi: 10.1080/03052150903386674

[13] Jun Zheng, Xinyu Shao, Liang Gao, Ping Jiang and Zilong Li. A hybrid variable-fidelity global approximation modeling method combining tuned radial basis function base and kriging correction, Journal of Engineering Design, 2013,Vol. 24, No. 8, 604–622.

[14] Benjamin Rosenbaum and Volker Schulz. Comparing sampling strategies for aerodynamic Kriging surrogate models, ZAMM ·Z. Angew. Math. Mech. 92, No. 11 – 12, 852 – 868 (2012)

[15] Mardia, K.V. and Marshall, R.J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71, 135–146.

[16] Hongbing Fang & Mark F. Horstemeyer (2006) Global response approximation with radial basis functions, Engineering Optimization, 38:04, 407-424, DOI:10.1080/03052150500422294

[17] Powell, M.J.D. Radial basis functions for multivariable interpolation: a review, in Proceedings of the IMA Conference on Algorithms for the Approximation of Functions and Data, (Mason, J. C. and Cox,M. G. (eds.), Oxford University Press, London, pp. 143–167.

[18] Queipo, N., R. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevintucker. 2005. “Surrogate-Based Analysis and Optimization.” Progress in Aerospace Sciences 41 (1): 1–28.

[19] Ai Yi Si. Research on Response Surface Method Optimization based on Radial Basis Function. Huazhong University of Science and Technology , 2012.

[20] Baker, C.A. et al, High-Speed Civil Transport Design Space Exploration Using Aerodynamic Response Surface Approximations, Journal of Aircraft, Vol. 39, No.2, 2002, pp. 215-220.

[21] Chun Y. Tang and Ken Gee. Generation of Aerodynamic Data using a Design of Experiment and Data Fusion Approach, 43rd AIAA Aerospace Sciences Meeting and Exhibit.2005. AIAA 2005-1137.

[22] Tang, C.Y. et al. Generation of Aerodynamic Data using a Design of Experiment and Data Fusion Approach[J]. AIAA 2005-1137

[23] Shan, S., and G. G. Wang. 2009. Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-Box Functions. Structural and Multidisciplinary Optimization. 41 (2): 219–241. doi: 10.1007/s00158-009-0420-2