



## Surrogate-Based Aerodynamic Design Optimization: Use of Surrogates in Aerodynamic Design Optimization

M. Y. M. Ahmed\* and N. Qin\*\*

**Abstract:** The role of optimization in various practical applications has been increasing over the years. In the field of aerodynamics, design optimization is rapidly evolving in many ways. One of the latest developments in this field is the introduction of surrogates or metamodels. The use of surrogates has a number of advantages especially concerning the computation cost, memory and time budgets. The present paper aims to introduce the topic of surrogate-based optimization. An overview on the main aspects of surrogates and the associated terminology are discussed in detail. In addition, the paper provides a survey over the previous efforts contributing in this field. The paper is finalized with some general recommendations.

**Keywords:** aerodynamic design optimization, computational fluid dynamics, surrogate modeling,

### I. Introduction

The application of optimization in the field of aerodynamics design is progressively increasing. Aerodynamic design optimization (*ADO*) is gaining an increasing interest motivated by the demands of handling sophisticated geometries, tackling more realistic flight conditions, and satisfying increasing design requirements (objectives). Since almost all modern aerodynamic design activities rely on numerical simulation computer codes (flow solvers), the use of *ADO* was strongly encouraged by the increasing capabilities of modern computers and the continuously improving numerical schemes, simulation, and optimization algorithms.

Nevertheless, the direct coupling of numerical simulation codes with search algorithms shows a number of drawbacks. Surrogate models (also known as metamodels) are introduced as an alternative that can remedy much of these drawbacks. Surrogates replace the simulation codes in the sense that the search is directly coupled with them rather than the simulation codes.

---

\* PhD Student, Department of Mechanical Engineering, University of Sheffield, [m.ahmed@sheffield.ac.uk](mailto:m.ahmed@sheffield.ac.uk)

\*\* Professor of Aerodynamics and Fluid Mechanics, Department of Mechanical Engineering, University of Sheffield, Senior Member AIAA, [n.qin@sheffield.ac.uk](mailto:n.qin@sheffield.ac.uk)

This paper is devoted to the following:

- illustrating the concept of surrogates, their types, and construction, and
- surveying the related previous efforts in the field.

The remainder of this paper is organized as follows: in the next section, the main elements of surrogate modeling are discussed. Next, the main types of surrogates are discussed and compared. A closely related topic which is the design of experiment (DoE) is then discussed. A survey over the efforts contributing in this field is presented near the end of the paper and the paper finalizes with the main conclusions and general recommendations. All abbreviations mentioned throughout the paper are listed in the appendix.

## II. Use of Surrogate Models in Aerodynamic Design Optimization

In aerodynamic optimization problems, the flow fields are simulated using flow solvers based on computational fluid dynamics (*CFD*) techniques. These high-fidelity numerical simulation computer codes proved to be a reliable, efficient, flexible, and relatively cheap means of analysis and design especially compared with experimental methods. The main drawback of using numerical analysis codes is that they are computationally expensive, highly memory demanding, and time consuming. Despite the fact that most of the analysis codes already use simplified forms of the governing equations and the continuously increasing processing speeds and memory capacities of modern computers made the numerical analysis more manageable; the demand for more accurate real calculations and the search for more complicated designs arise as well.

These drawbacks of analysis codes become more severe when they are utilized in the field of shape optimization since it requires rather more computations. The conventional optimization methods that use gradient-based techniques or even non-gradient search techniques require hundreds or thousands of analysis code implementation. Consequently, the cost of performing optimization for complex designs becomes rather expensive with the possibilities of multiple local optima. One way to minimize the time scale of such operations is the implementation of parallel computing, thus handling more one than design at a time. However, this way is not applicable with gradient-based optimization algorithms since they deal with one design at a time. Another problem related to using the optimization methods in direct conjunction with high fidelity analysis codes which is the occurrence of numerical noise.

Typical sources of numerical noise include:

- The discrete representation of continuous governing equations and physical quantities,
- The numerical round-off errors, and
- The incomplete iterative convergence of calculations.

These errors can mislead the optimization algorithm or even cause local search methods to find fake optima. Another feature related to computer analysis codes is that a majority of them are originally designed as stand-alone modules and complications arise when trying to *interfere* their performance, in some cases the source codes are not accessible at all. Consequently, an increasing effort has been devoted to search for cheap alternatives of the high fidelity analysis codes in analysis, design, and optimization. One of these alternatives is the use of approximation models, metamodels, or surrogates.

The basic idea of the using the surrogates (approximation models) is to *replace* the high fidelity, expensive analysis code with a less expensive *approximate* model. The high-fidelity model is commonly represented by the functional relation

$$y = f(\mathbf{x}) \quad (1)$$

where  $\mathbf{x}$  is the vector of design parameters (inputs) and  $y$  is the output (objective value). On the other hand, a surrogate model is expressed as

$$\hat{y} = \hat{f}(\mathbf{x}, \boldsymbol{\alpha}) \approx f(\mathbf{x}) \quad (2)$$

which is intended to be less computationally expensive than the high fidelity code. Here  $\boldsymbol{\alpha}$  is a vector of undetermined parameters that must be evaluated prior to applying the surrogate.

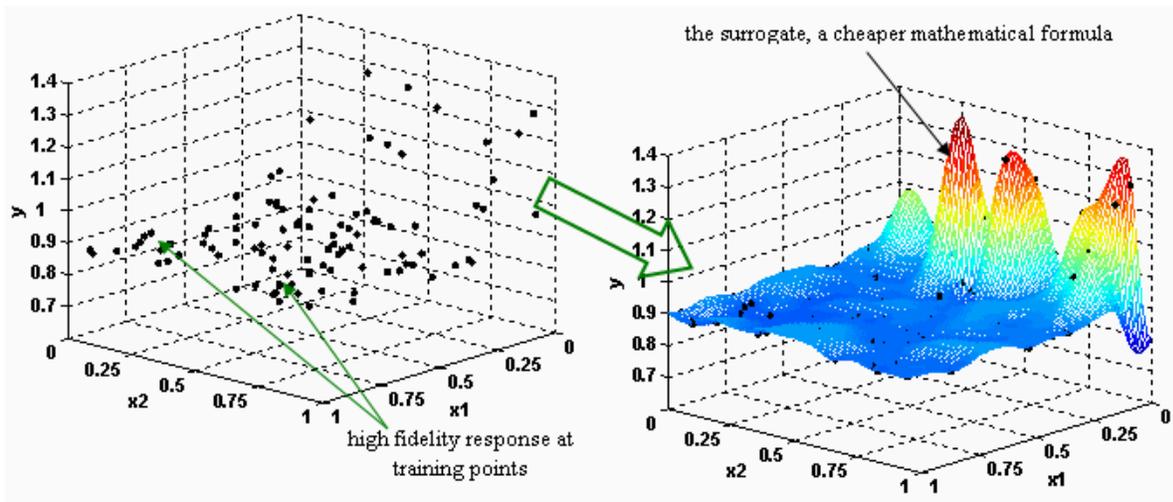
### II.1. Classification of Surrogate Models

Surrogates can be classified into two main categories according to their approximation strategy, namely, the black-box and physics-based approaches. In the black-box approach, the high fidelity analysis code is considered a stand-alone module that can not be modified; hence the name. Only the inputs and output(s) of this black box are of concern. Consequently, the approximation model uses a set of points to build a less-expensive model that "approximates" the real, more expensive inputs-output(s) relation.

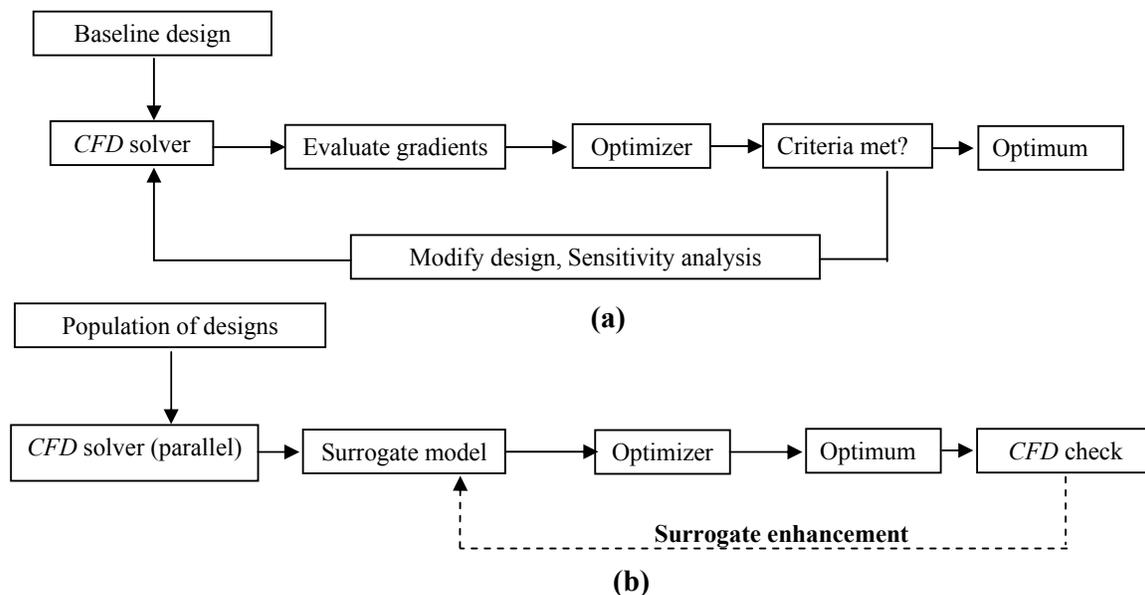
In contrast, the physics-based approach exploits to some extent the analysis code. It introduces modifications to simplify the governing equations either in their continuous or discrete forms and eventually, make them computationally less expensive. Obviously, this approach requires access and a good understanding of analysis code structure. In this study, we only focus on the black-box approach.

Generally, black-box models use a set of  $n$  selected design points,  $x$ , called the training set from the whole design space,  $\Omega$ ; the concept of this selection will be discussed later. The high fidelity code is then applied to each member of the training set to evaluate the respective outputs,  $y$ , possibly making use of parallel computation facilities. The unknown parameters  $\boldsymbol{\alpha}$  are then estimated and the approximation model,  $\hat{y}$ , is eventually constructed. Hopefully, the surrogate manages to accommodate the training points and accurately represent the real input-output relation such that it can predict the response at new points. Fig. 1 illustrates how a surrogate replaces the high fidelity response at training points.

The surrogate is then examined to test its accuracy generally using new points other than the training set. Finally, the surrogate is used in lieu of the high-fidelity code in design and optimization. The high fidelity model is overlooked and the subsequent search is made directly on the surrogate model. An additional step is required in surrogate-based optimization; the optimum design attained by the search technique must be validated using the high fidelity model. In some cases, the surrogate model accuracy is improved by augmenting the training set with the optimum; this can be repeated until no further improvement is achieved. The role of surrogates in aerodynamic design optimization can be visualized in Fig. 2.



**Fig. 1. A Good Surrogate Incorporates the Response at all Training Points**



**Fig. 2. Basic Elements of (a) CFD-based and (b) CFD-surrogate-based ADO**

Despite that the computation cost of constructing the surrogate may be significant; this cost is incurred prior to the use of the surrogate in design and optimization. Consequently, the surrogate can be evaluated hundreds and thousands of times without significant computation cost. However, the mathematical operations involved in constructing and implementing some metamodels may become computationally expensive as the number of design parameters and design points increases that the relative superiority of using metamodels starts to diminish; which is known as the *curse of dimensionality*. Hence, a good selection of training points and a neat setup of the *proper* surrogate are crucial to yield a reliable substitute of the high fidelity flow solver.

The use of surrogate models has a number of appealing advantages over the direct use of analysis codes in design. These advantages include:

- Less computation cost.
- Smoothing out the numerical noise arising from numerical simulation.
- No modification to the analysis code is needed.
- Surrogate models give better "feeling" and understanding of the physical inputs-outputs relations.

When used in design optimization, surrogates give additional benefits such as:

- Smoothing out the noise arising from evaluating the objective and constraint functions.
- Easy adaptation to parallel and distributed computations.
- Separating the analysis code from the optimization algorithms.
- Easy integration of various codes used in multidisciplinary optimization.

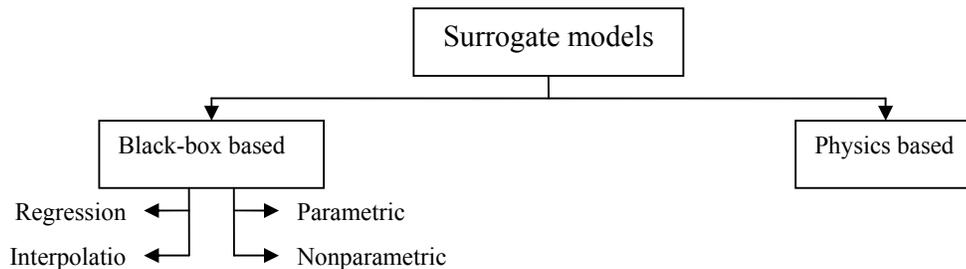
Basically, the approximation methods that adopt the black-box approach can be classified according to two main criteria as follows:

- 1- According to the nature of the unknown parameters,  $\alpha$ ; surrogate models can be either parametric or nonparametric.
  - Parametric models use the initial training set to estimate the unknown parameters  $\alpha$ ; once they are known, the training set are no longer used in making predictions at new points and only the parameters, that are already known, decide the response at the new points. Polynomial regression (*PR*) and multivariate adaptive regression splines (*MARS*) are examples of parametric methods.
  - Nonparametric models use the initial training set to estimate the unknown parameters  $\alpha$  and continue using the training set in predicting new design points. The response at the new points depends on both the parameters and all the training set points. Kriging (*KG*), artificial neural networks (*ANN*), radial basis functions (*RBF*), and support vector machines (*SVM*) are examples of nonparametric methods.
- 2- According to the technique of building the surrogate model; surrogates can use either interpolation or regression techniques. It is worth explaining first the main differences between physical (real) and numerical (computer) experiments.

Physical vs. numerical experiments: The major difference between physical and numerical experiments is the nature of error. The errors in a physical experiment are of a random nature; the values of errors change even if the same experiment is perfectly repeated. The error "noise" is assumed to be identically and independently distributed. In contrast, errors in numerical experiments are of a deterministic nature; the values of errors will remain exactly the same whenever the same numerical experiment is repeated with the same inputs.

- Regression models are primarily designed for physical experiments. Nevertheless, they are widely used with numerical experiments. In this type, a single function (usually polynomial) is designed to incorporate all the available points in the training set. An error term is added to the model to account for the independently and identically distributed random noise. Polynomial regression (*PR*), multivariate adaptive regression splines (*MARS*), artificial neural networks (*ANN*), and support vector machines (*SVM*) are regression models.

- Interpolation models are more appropriate for computer experiments. In this type, a simple base function (generally low order polynomials) interpolating the training set points is added to another function representing the local deviations at each point. The error noise is treated as a deterministic error that has fixed values at the respective design points. Kriging (*KG*), and radial basis functions (*RBF*) are interpolation models. Fig. 3 summarizes the various classes of surrogate models



**Fig. 3. Classification of Surrogate Models**

## II.2. Basic Steps of Surrogate Modeling

Generally speaking, surrogate modeling is an iterative procedure that involves the main following steps:

- Data generation,
- Model-structure selection,
- Parameter estimation, and
- Model validation.

### II.2.1. Data Generation, Data Building Techniques

The appropriate selection of the number and location of the training set points is crucial for designing a good surrogate model. In fact, the number of points in the training set increases with the accuracy needed in the model, the number of design parameters, and the degree of nonlinearity in the input-output functional relation. Data building is faced with two conflicting targets:

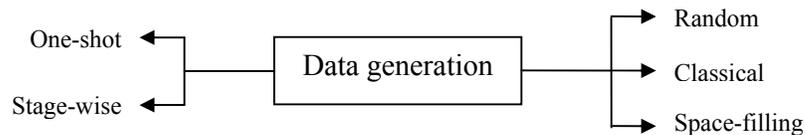
- Increasing the model accuracy; by incorporating as large training set as possible.
- Decreasing the computational cost of data generation by selecting as few training points as possible since all these points are evaluated by the expensive high-fidelity analysis code.

Selection depends on the nature of experiment (physical or numerical), geometry of design space (regular or irregular), and target of surrogate modeling (analysis, design, or optimization). Nevertheless, the selection techniques can be classified according to two criteria:

- 1- According to location of selected points: Points can either have a random, classical, or space-filling fashion. Sample selection is explained in more detail later.
- 2- According to the building strategy: training set can be generated once (one-shot) or sequentially (stage-wise).

- In the one-shot strategy, all point in the training set are selected and evaluated by the high-fidelity code at the beginning of the surrogate construction process; no points are added during this iterative process. It is a relatively easy technique but it is less flexible and can be very computationally expensive in designing the surrogate model.
- The stage-wise strategy continuously adds new points to the training set during the iterative process of surrogate design. It involves more complicated steps but it gives more flexibility and continuously incorporates the nonlinearity in input-output relation.

Fig. 4 summarizes the various strategies in data generation



**Fig. 4. Classification of Data Generation Techniques**

### II.2.2. Model Structure Selection

Once the training points are selected and evaluated by the computationally expensive high-fidelity analysis code, they are manipulated to construct the less expensive surrogate model. As stated earlier, the model can be parametric or nonparametric, regression or interpolation. Parametric models are relatively easier to implement whereas nonparametric models have superior utility when applied to cases with complicated input-output relations.

Regression models are relatively easier to develop and implement. On the other hand, the computation cost increases when dealing with more complicated input-output relations. This is because more points must be added to the training set and higher order polynomials must be used with possible instabilities and false optima. Interpolation models are more appropriate for computer experiments. Interpolation models are relatively more complicated in construction and implementation but they are more flexible and cope better with complicated input-output relations.

### II.2.3. Parameter Estimation

After selecting the surrogate model that is believed to accurately simulates and input-output relation, the undetermined parameters,  $\alpha$ , must be estimated. The concept of estimation is known as the "maximum likelihood" and is applied by minimizing a suitable error function. In its simplest form, assuming Gaussian error, the maximum likelihood reduces to the least square error function. There are many variants for the error function minimization.

### II.2.4. Model Validation

After fully building the surrogate model, a check for its suitability for the case in concern must be made. Basic "goodness-of-fit" tests such as maximum error, root mean square error, and sum of errors are the typical diagnostic checks.

### II.3. Use of Surrogate Models in Design Optimization

The strategies of incorporating surrogates within optimization methodologies are often referred to approximation model management frameworks (*AMMF*), model management frameworks (*MMF*), or surrogate management frameworks (*SMF*). These strategies vary depending on the surrogate type, optimization algorithm, and target of optimization. Their main strategies are: trust-region, space mapping, the use of global surrogate models, and the use of surrogates with evolutionary algorithms. They are briefly explained below.

#### Trust Region Methods

In this strategy, gradient-based optimizers such as steepest descent are applied to the surrogate function. The surrogate gradient is evaluated and the step size (also called the move limit) is chosen to ensure improvements in the surrogate response. Implementation of this technique requires gradient information from the high fidelity analysis and internal optimization to evaluate the move limit. It can be used for local optimization and design improvement purposes.

#### Space Mapping

The basic idea in this strategy is to use variable fidelity models. A low-fidelity (coarse) model is used to "map" the whole design space. Next, the model is optimized to locate the optimum points. A high-fidelity (fine) model is used to evaluate the response at the "promising" points and they are added to the data set. No gradient information is needed in this method but some complicated complications may require additional schemes and the method is suitable for local optimization.

#### The Use of Global Models

In this technique, an initial set of design points is created and the high-fidelity model is used to evaluate the response at each point. Then, the surrogate model is constructed and the promising points are located. The response at these specific points is evaluated by the high-fidelity model and the points are added to the dataset. The process continues until the convergence criteria are satisfied. This basic key point in this method is the proper management of the design space sampling. Basically, the initial set must be space-filling to guarantee a surrogate that accurately simulates the input-output relation. On the other hand, it is demanded by the optimization algorithm to place more points near the promising points and regions with high curvature. Balance should be made between the two conflicting targets. Nevertheless, this conflict can be met using some techniques such as the expected improvement (*EI*) which aims to locate the new points based on some optimization criterion.

#### The Use of Surrogates with Evolutionary Algorithms

Here, the surrogate model is used in conjunction with evolutionary algorithms (*EAs*) such as genetic algorithms. The initial design data set is selected and the high-fidelity model is used to evaluate the corresponding response. A surrogate model is constructed and used to evaluate the response (fitness). The standard *EA* operators are applied to generate new population and the data set is updated. The process continues until satisfaction criteria are met.

### III. Main Types of Surrogate Models

In the literature, we find a variety of surrogates. However, the following metamodels are found to be the main models used in design and optimization applications.

#### III.1. Polynomial Regression (PR)

Polynomial regression is a parametric regression metamodel; also known as response surface model, *RSM*. It was originally tailored for physical experiments where the output is characterized by random errors [1]. In this method, the points in the training set are fit by a polynomial plus an error.

$$\hat{y} = y + \varepsilon \quad (3)$$

where  $\varepsilon$  represents the random error which is assumed to be normally distributed with zero mean and variance  $\sigma^2$ . The approximation polynomial can have any order; however, it is typically of the first order (linear) or second order (quadratic) ones,

$$\hat{y} = \beta_o + \sum_{i=1}^k \beta_i x_i \quad (4)$$

$$\hat{y} = \beta_o + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \sum_{j \geq i}^k \beta_{ij} x_i x_j \quad (5)$$

respectively, where  $\beta_o$  is called the intercept,  $\beta_{ij}, i \neq j$  are the interaction coefficients,  $x_i$  refers to one of the  $k$  design parameters. The polynomial coefficients  $\beta$ 's (also known as the regression parameters) represent the undetermined parameters vector  $\boldsymbol{\alpha}$ . In the quadratic case, the total number of the undetermined parameters is  $m = (k+1)(k+2)/2$ . Parameters are estimated via minimizing the sum of squares of the deviations of predicted values,  $\hat{y}(\mathbf{x})$ , from the actual high-fidelity values,  $\mathbf{y}(\mathbf{x})$ , using the equation:

$$\boldsymbol{\beta} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{y} \quad (6)$$

where  $n$  is the number of training set points and  $\mathbf{y}$  is a column vector that contains the exact values of the response at all training point estimated by the high fidelity model.  $\mathbf{A}^T$  is the transpose of matrix  $\mathbf{A}$  which is an  $n \times m$  matrix of training set points expressed as:

$$\mathbf{A}_{(n \times m)} = \begin{bmatrix} \mathbf{1} & x_{11} & \dots & x_{11}x_{12} \dots & x_{1k}^2 \\ \mathbf{1} & x_{21} & \dots & \dots & x_{2k}^2 \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{1} & x_{n1} & \dots & x_{n1}x_{n2} \dots & x_{nk}^2 \end{bmatrix} \quad (7)$$

The uncertainty in the computed parameters is evaluated using the  $t$ -statistic. The  $t$ -statistic for the  $j$ th parameter  $\beta_j$  is expressed as:

$$t = \frac{\beta_j}{\sqrt{\hat{\sigma}^2 C_j}} \quad (8)$$

where  $C_j$  is the  $j$ th diagonal element of the matrix  $(\mathbf{A}^T \mathbf{A})^{-1}$  and  $\hat{\sigma}^2$  is the error variance defined as

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{A}^T \mathbf{y}}{n - m} \quad (9)$$

A low value of  $t$ -statistic indicates that the corresponding parameter was not accurately estimated and the overall accuracy of the model can be improved by dropping such parameters [1]. The response at a new point  $\bar{\mathbf{x}}$  in the design space,  $\hat{y}(\bar{\mathbf{x}})$ , is directly evaluated by substituting for  $\bar{\mathbf{x}}$  in the polynomial equation.

### III.2. Kriging (KG)

Kriging, also known as design and analysis of computer experiments *DACE*, is a nonparametric interpolation model based on Gaussian stochastic process models. In contrast to *RSM*, kriging was originally tailored for computer experiments characterized by deterministic errors [2]. Kriging imposes a global model (functional relation) that interpolates all design points plus "localized" functions representing the deviations (departure) from the global model at all points. The response is expressed in terms of the design parameters as

$$\hat{y}(\mathbf{x}) = f(\mathbf{x}) + z(\mathbf{x}) \quad (10)$$

where  $f(\mathbf{x})$  is a low-order polynomial that interpolates the design points. Typically, a constant value was found sufficient for modeling complex input-output relations [3]. Hence, the output can be viewed as a random field with mean  $\boldsymbol{\beta}$ ,

$$\hat{y}(\mathbf{x}) = \boldsymbol{\beta} + z(\mathbf{x}) \quad (11)$$

$z(\mathbf{x})$  is a Gaussian stochastic function that represents the realization of random process with zero mean, variance  $\sigma^2$ , and covariance given by

$$\text{Cov}(Z) = \sigma^2 \mathbf{R}(\mathbf{x}^i, \mathbf{x}^j) \quad (12)$$

where  $\mathbf{R}(\mathbf{x}^i, \mathbf{x}^j)$  is the correlation matrix which is an  $n \times n$  symmetric matrix with ones in the diagonal and  $n$  is the number of training set. Other terms in the matrix are given by the spatial correlation function

$$R_{ij} = \prod_{l=1}^k \text{scf}_k(x_k^i - x_k^j) \quad (13)$$

where  $i$  and  $j$  denote two training points,  $l$  refers to a design parameter, and  $k$  is the number of design parameters. The spatial correlation function can be viewed as functions of the "weighted" distance between samples. It can have many forms; typically a linear, a cubic, or an exponential (Gaussian) function

$$scf(d) = \begin{cases} 1 - \frac{1}{\theta}|d|, & |d| < \theta \\ \mathbf{0} & |d| \geq \theta \end{cases} \quad (14)$$

$$scf(d) = \begin{cases} 1 - 6\left(\frac{d}{\theta}\right)^2 + 6\left(\frac{|d|}{\theta}\right)^3, & |d| < \frac{\theta}{2} \\ 2\left(1 - \frac{|d|}{\theta}\right)^3, & \frac{\theta}{2} \leq |d| < \theta \\ \mathbf{0}, & |d| \geq \theta \end{cases} \quad (15)$$

$$scf(d) = \exp(-\theta \times d^p) \quad (16)$$

The mean parameter,  $\boldsymbol{\beta}$ , is evaluated by minimizing the sum of squares of error of the estimated error using the equation

$$\boldsymbol{\beta} = [\mathbf{A}^T \mathbf{R}^{-1} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{R}^{-1} \mathbf{y} \quad (17)$$

where  $\mathbf{A}$  is an  $n \times m$  matrix of training set points depending on the choice of the function  $f(\mathbf{x})$ ; in the constant case,  $\mathbf{A}$  is a  $n \times 1$  vector of all ones. The parameters  $\theta$  (and  $p$ ) that ensure the "best fit" of the model to the training data are evaluated by maximizing the likelihood estimation *MLE* (or minimizing its negative). *MLE* can have a variety of forms:

$$-\frac{1}{2} \left[ n \ln(2\pi) + n \ln \sigma^2 + \ln |\mathbf{R}| + \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{A}\boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\beta}) \right] \quad (18)$$

$$\frac{1}{(2\pi\sigma^2)^{n/2} \sqrt{|\mathbf{R}|}} \exp \left[ \frac{-(\mathbf{y} - \mathbf{A}\boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\beta})}{2\sigma^2} \right] \quad (19)$$

$$-\frac{1}{2} \left[ n \ln(2\pi) + n \ln \sigma^2 + \ln |\mathbf{R}| + n \right] \quad (20)$$

$$-\frac{1}{2} \left[ n \ln \sigma^2 + \ln |\mathbf{R}| \right] \quad (21)$$

$$-\frac{n}{2} \left[ \ln \sigma^2 + \ln |\mathbf{R}| \right] \quad (22)$$

where the maximum likelihood estimation of  $\sigma^2$  is expressed as:

$$\sigma^2 = \frac{1}{n} (\mathbf{y} - \mathbf{A}\boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\beta}) \quad (23)$$

Generally, evaluating the *MLE* is a nonlinear unconstrained optimization sub-problem where  $\sigma^2$  and  $\mathbf{R}$  are functions of the unknown parameters. The response at a new point  $\bar{\mathbf{x}}$  in the design space,  $\hat{y}(\bar{\mathbf{x}})$ , is directly evaluated by applying the equation

$$\hat{y}(\bar{\mathbf{x}}) = \boldsymbol{\beta} + \mathbf{r}^T(\bar{\mathbf{x}}) \mathbf{R}^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\beta}) \quad (24)$$

where  $\mathbf{r}(\bar{\mathbf{x}})$  is the correlation vector between  $\bar{\mathbf{x}}$  and all points in the training set.

### III.3. Radial Basis Functions (RBF)

Radial basis functions method is a nonparametric regression modeling technique. It is designed to suit outputs with deterministic errors. The method uses linear combinations of radially symmetric functions based on the Euclidean distance from a given "centre" as the basis functions to approximate response functions. A radial basis function model can be expressed as

$$\hat{y} = \sum_{i=1}^n \alpha_i K(\|\mathbf{x} - \mathbf{x}^i\|) \quad (25)$$

where  $i$  refers to one of the  $n$  training points which is the centre of the function  $K$  and  $\alpha$  is the corresponding weight factor. The radial function  $K$  can have many forms, typically, linear, cubic splines, Gaussian, multi-quadratic, and inverse multi-quadratic

$$K = \|\mathbf{x} - \mathbf{x}_c\| \quad (26)$$

$$K = \|\mathbf{x} - \mathbf{x}_c\|^3 \quad (27)$$

$$K = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_c\|^2}{\theta}\right) \quad (28)$$

$$K = \left(1 + \frac{\|\mathbf{x} - \mathbf{x}_c\|^2}{\theta}\right)^{0.5} \quad (29)$$

$$K = \left(1 + \frac{\|\mathbf{x} - \mathbf{x}_c\|^2}{\theta}\right)^{-0.5} \quad (30)$$

respectively, where  $\theta$  is called the "shape parameter" of the function. It controls the domain of influence of the respective function and is chosen according to the input-output relation. The set of weights  $\boldsymbol{\alpha}$  represents the undetermined parameters in this method. They are evaluated by solving the algebraic system of equations

$$\mathbf{K}\boldsymbol{\alpha} = \mathbf{y} \quad (31)$$

where  $\mathbf{K}$  is an  $n \times n$  symmetric matrix formed by the training points such that

$$K_{ij} = K(\|\mathbf{x}^i - \mathbf{x}^j\|) \quad (32)$$

$\alpha$  is an  $n \times 1$  vector of weights, and  $\mathbf{y}$  is an  $n \times 1$  vector of outputs at the training points. The response at a new point  $\bar{\mathbf{x}}$  in the design space,  $\hat{y}(\bar{\mathbf{x}})$ , is directly evaluated by applying the equation

$$\hat{y}(\bar{\mathbf{x}}) = \sum_{i=1}^n \alpha_i K(\|\bar{\mathbf{x}} - \mathbf{x}^i\|) \quad (33)$$

#### III.4. Multiple Adaptive Regression Splines (MARS)

Multiple adaptive regression splines method is a parametric regression method. In *MARS*, the range of variation of each input (design parameter) is divided into sub-intervals. The locations at the start and end of each sub-interval are termed "knots". *MARS* aims to approximate input-output functional relations by adaptively selecting a set of basis functions (splines). This is done through a forward/backward iterative approach on the different intervals of the independent parameter ranges. The *MARS* surrogate model can be expressed as

$$\hat{y} = \sum_{m=1}^M a_m B_m \quad (34)$$

where  $a_m$  is the undetermined regression parameter and  $M$  is the total number of basis functions,  $B_m$ , which can be expressed as

$$B_m(x) = \prod_{i=1}^k [s_{i,m}(x_{i,m} - t_{i,m})]_+^q \quad (35)$$

where  $x_{i,m}$  is the  $i$ -th input of the  $k$  of design parameters,  $t_{i,m}$  is the knot location of the corresponding parameter,  $s_{i,m}$  equals  $\pm 1$ ,  $q$  is the power of the function, and the  $+$  subscript indicates that the power function is truncated (assigned to zero outside the knots) as

$$[s_{i,m}(x_{i,m} - t_{i,m})]_+^q = 0 \quad , \quad s_{i,m}(x_{i,m} - t_{i,m}) \leq 0 \quad (36)$$

That key parameter in this model is to specify the location of knots which is dictated by the type of input-output relation and shape of design space. The undermined regression parameters are evaluated by minimizing the sum of error squares.

#### III.5. Artificial Neural Networks (ANN)

Artificial neural network model is a nonparametric regression method. This method utilizes the functional concept of neurons in the brain. A neural network is composed of neurons which can be represented as nonlinear transfer functions of the inputs. A typical configuration of the *ANN* surrogate is the feed-forward neural network which can be expressed as:

$$\hat{y} = \sum_{i=1}^m \alpha_i \phi(a_i) \quad (37)$$

$$\text{and } a_i = \sum_{j=1}^k \gamma_{ij} x_j + \beta_j \quad (38)$$

where  $\alpha$ ,  $\gamma$ , and  $\beta$  are the undetermined parameters,  $k$  is the number of input parameters, and  $\phi$  is a transfer function, typically on the form

$$\phi(a) = (1 + e^{-a})^{-1} \quad (39)$$

### III.6. Models Validation

In the literature, we find many techniques for testing the fitness of the surrogate; some of them are unique to certain types of surrogates. For example, regression error is used with *PR* [4], sample sensitivity and cross validation are used with kriging and *RBF* [3, 5]. Nevertheless, there are a number of *universal* techniques that are used to assess and compare the accuracy of the various surrogated. Of these techniques, the error variance of the training set, also called the process variance, is used to assess the fitness of the model with the training points [3, 6].

However, the fitness of the surrogate is generally assessed using new points other than the training ones. The surrogate performance in predicting these *untried* points is tested using a number of methods mainly: the root mean square, *RMS* [7, 8], the maximum absolute, *MAE* [7], and the  $R^2$  [7- 9] error measures.

$$RMS = \sqrt{\frac{1}{q} \sum_q (y_i - \hat{y})^2} \quad (40)$$

$$MAE = \text{Max}_q (|y_i - \hat{y}_i|) \quad (41)$$

$$R^2 = 1 - \left[ \frac{\sum_q (y_i - \hat{y}_i)^2}{\sum_q (y_i - \bar{y})^2} \right] \quad (42)$$

where  $\bar{y}$  is the mean of the exact response at the  $p$  untried points  $y_i$  and  $\hat{y}$  is the corresponding surrogate response value. A smaller value of *RMS* and *MAE* and a higher value of  $R^2$  indicate better surrogate accuracy.

### III.7. Comparison of the Various Surrogate Models

Having discussed the basic metamodels, it is important at this point to hold some comparisons to get a feel of the pros and cons and the range of application of each. In fact, we can find some comparisons in the literatures. We can summarize all these efforts in Table 1. For the interested reader, we recommended [3, 10, 11] for more details.

**Table 1. Basic Features of the Main Surrogate Models**

Features Models	Advantages	Disadvantages
Polynomial Regression (PR)	<ul style="list-style-type: none"> <li>- well-established model, easy to implement, require fewer calls of the high fidelity model.</li> <li>- better performance with low order non-linear functions</li> <li>- best suited for small scale applications with less than 10 design parameters.</li> <li>- due to the "smoothing" capabilities of polynomials, optimization converges relatively faster with noisy functions</li> </ul>	<ul style="list-style-type: none"> <li>- a drawback is expected with highly non-linear and irregular performance problems.</li> <li>- higher order polynomials can be adopted. However, a large number of training points is needed. Using higher-order polynomials invokes instabilities and yield false optima.</li> </ul>
Kriging (KG)	<ul style="list-style-type: none"> <li>- well suited for numerical experiments with deterministic errors.</li> <li>- extremely flexible by virtue of a wide range of correlation functions.</li> <li>- even when a constant term is used for the global part of the model, the performance is comparable to a quadratic regression model.</li> <li>- has a higher performance when applied to low-order nonlinear and large scale problems over a wide range of samples size and design</li> </ul>	<ul style="list-style-type: none"> <li>- model construction can be time consuming for large scale problems. It involves matrices operations, optimization sub-problems. Such costs may overweight the benefits.</li> <li>- if the training points are relatively "close" to each other, the correlation matrix can become singular.</li> <li>- Additional points are needed to assess the accuracy of the model.</li> </ul>
Radial Basis Functions (RBF)	<ul style="list-style-type: none"> <li>- yields good results for a wide range of sampling size and design. Produces good approximations to response functions with various patterns of both random and deterministic errors.</li> <li>- for high-order nonlinear and small scale problems, RBF models are relatively more accurate and robust</li> </ul>	<p>the evaluation of the model parameters involves computational complications especially with large number of design parameters; such computations can be highly expensive regarding computation and memory.</p>
Multiple Adaptive Regression Splines (MARS)	<ul style="list-style-type: none"> <li>- relatively more accurate and flexible compared with the standard PR since the design space is split into sub-intervals rather than a single one.</li> <li>- computation cost of model construction is significantly reduced.</li> </ul>	<ul style="list-style-type: none"> <li>- a relatively new technique, more assessment studies are required.</li> <li>- since MARS is based on dividing the design space, it fails for small training set size. Caution must be paid to selecting the knot locations depending on problem nature and designer experience.</li> </ul>
Artificial Neural Networks (ANN)	<p>yield better approximations compared to the classical response surface methods in cases if the nature of the problem is unknown, involves large number of design parameters, or not completely bounded design spaces</p>	<ul style="list-style-type: none"> <li>- a relatively large number of parameters are involved; the evaluation of these parameters requires high computation and memory requirements.</li> <li>- many sophisticated procedures for estimating the number of neurons, initialization, training, and regularization.</li> </ul>

## IV. Design of Experiments; Sampling Techniques

Sampling, known also as design of experiment *DoE*, is a governing step in the construction of surrogates; a good sampling design will generally improve the performance of the surrogate model. In this section, the term "factors" will be used to refer to the design parameters and the term "levels" will refer to the finite number of possible value each factor can take. Next we explain how data can be prepared prior to sampling. Later we present a brief survey over the various sampling techniques.

To prepare the design parameters for sampling, they should be on the form

$$x_i^l \leq x_i \leq x_i^u, \quad i \in [1, k] \quad (43)$$

where  $x_i$  is the  $i$ -th factor, and  $u$  and  $l$  denote the corresponding upper and lower bounds (given by the designer), respectively. Hence, the design space  $\Omega$  can be viewed as a hypercube of  $k$  dimensions bounded by the factors. A bounded unconstrained optimization problem implies a regular design space (cube) whereas constraints impose irregularities to the design space. Next, to eliminate the effect of different scales, all bounds are scaled to a common range, typically,  $[-1,1]$  or  $[0,1]$ . Consequently, all factors are normalized to lie within the range. In what follow, we will briefly describe specific features of most of the designs used in surrogate construction.

### IV.1. Random Designs

Random designs such as Monte Carlo are the simplest and most straightforward *DoE* technique. The points are randomly selected inside the design space. A random generator algorithm is involved in the process. The major disadvantage of the random technique is that the generated points are not space-filling and there is no guarantee that these points will be concentrated where significant variations and curvature of the response function exist.

### IV.2. Classical Designs

The classical techniques were primarily designed for physical experiments where random errors are expected. However, they are used in numerical experiments as well. The basic idea is to select points from the design space extremes. The classical technique is relatively simple and easily implemented. Full factorial design (*FFD*), partial factorial design (*PFD*), and face centered cubic (*FCC*) are examples of the classical technique. Such technique does not guarantee space-filling which is preferred for numerical experiments. In addition, the number of design points grows rapidly, sometimes exponentially, with factors and levels. They are also not suitable for irregular design spaces.

#### IV.2.1. Full Factorial Design (FFD)

It is the most basic classical design. The number of design points is dictated the number of levels and factors. The most common are the  $2^k$  full factorial in which there exist two levels for each factor, namely, at the lower and upper levels and the  $3^k$  full factorial in which there exists three levels for each factor; at the lower, upper, and mid levels, Fig. 5.a.

#### IV.2.2. Partial Factorial Design (PFD)

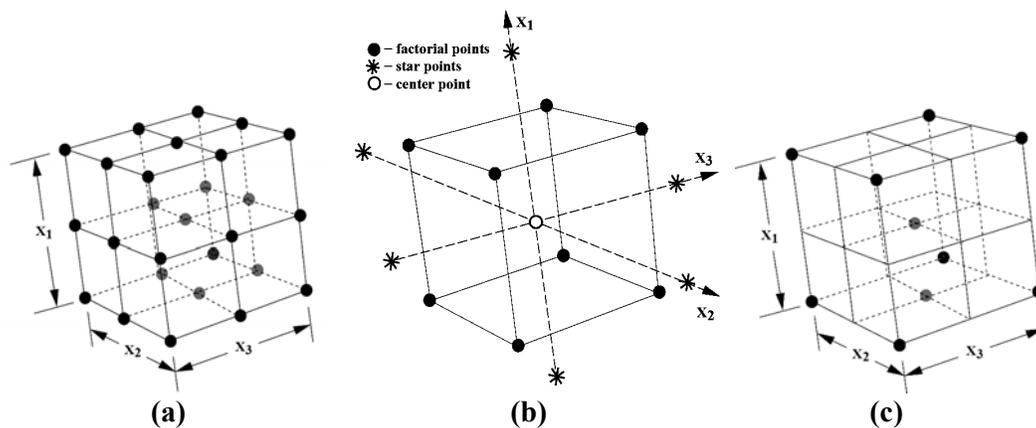
It is used in lieu of full factorial for large experiments. Only a fraction of the full factorial is used. The common designs are  $2^{k-1}$  and  $3^{k-1}$ . Consequently, it is relatively less expensive but rather less space-filling, too.

#### IV.2.3. Central Composite Design (CCD)

This design is a two-level (full or fractional) factorial design, augmented by a number of center points and two "star" points for each factor. The star points are positioned at a distance  $1 - \sqrt{k} \leq d \leq 1 + \sqrt{k}$  for the central point. A relatively better space-filling but higher cost characterize this design, Fig. 5.b.

#### IV.2.4. D-optimal

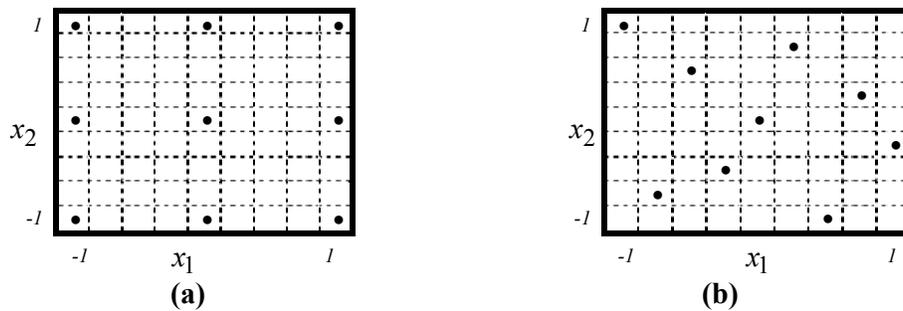
A D-optimal experimental design is a collection of sample sites in a design space for which the value of the determinant  $|\mathbf{X}^T \mathbf{X}|$  over all possible locations is maximized.  $\mathbf{X}$  is the  $n \times k$  design matrix which contains all design points,  $k$  is the number of factors, and  $n$  is the number of design points. A number of "candidate" design points are selected and an iterative numerical optimization method is employed to locate the best design points. D-optimal has many advantages over classical designs such as it can deal with irregular design spaces and being considerably less expensive. However, it is extremely expensive and time consuming especially for high-scale problems; more than 5 factors [12], Fig. 5.c.



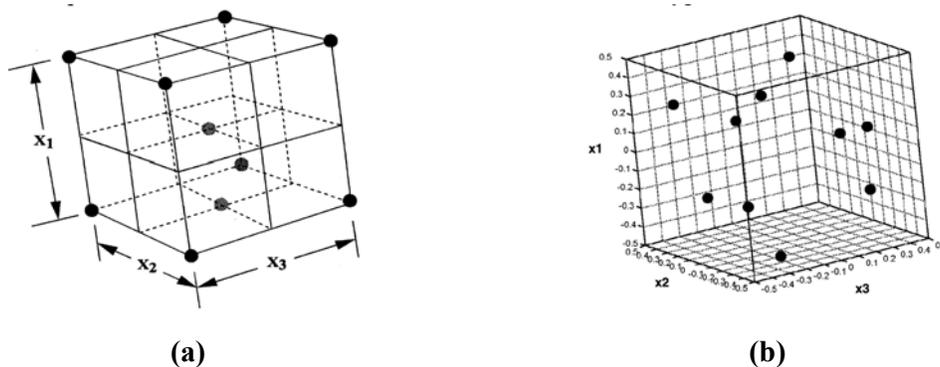
**Fig. 5 Graphical Representation of Classical Designs for a 3-factor, 3-level Problem**  
(a) FFD, 27 points, (b) CCD, 15 points, and (c) D-optimal, 10 points

### IV.3. Space Filling Designs

In the space-filling technique, points tend to cover uniformly the whole design space rather than focusing on the extremes only. Such technique is more suited for numerical experiments where deterministic errors are more likely to occur. Space filling designs are relatively more sophisticated in implementation. Orthogonal arrays (OA), Latin hypercube sampling (design) (LHS, LHD), and minimum discrepancy sequences adopt the space-filling technique. Figures 6 and 7 illustrate typical design spaces for two and three factors for classical and space filling techniques, respectively.



**Fig. 6 Typical Sampling of 2-factor, 9-point Space with (a) Classical and (b) Space-Filling Design**



**Fig. 17 Typical Sampling of 3-factor, 10-point Space with (a) D-optimal and (b) LHD**

#### IV.3.1. Latin Hypercube Design, Sampling (LHD, LHS)

The basic idea of the *LHD* is to divide each factor  $k$  into  $p$  levels of equal probability. Consequently, when a one-dimensional projection of the hypercube is taken, there will be a single sample point at each level. A Latin hypercube is a matrix of  $p$  rows and  $k$  columns. This design has the advantage of offering flexible sample sizes and well-distributed sampling. On the other hand, space-filling is not always guaranteed. Some modifications, e.g. optimal *LHD*, were introduced to ensure evenly distributed points.

#### IV.3.2. Uniform Designs

A uniform design is simply an *LHD* with more uniformity. Viewing the design space to be divided in "cells", the main difference between *LHD* and uniform design is that in uniform design, samples are situated at the center of each cell rather than randomly selected inside the cell. A uniform design adds uniformity to the design with no extra points (cost) added.

#### IV.3.3. Orthogonal Array (OA)

An orthogonal array is a matrix of  $p$  rows and  $k$  columns with every element being an integer between 0 and  $q-1$ . An orthogonal array has a strength  $t$  reflecting the number of combinations of  $l$  levels appearing in any of the  $r$  columns of the array. Orthogonal array can be viewed as an *LHD* with uniform spacing between points when projected to any factor.

#### IV.3.4. Minimum Discrepancy Sequences

The discrepancy of a design matrix is a measure of non-uniformity of samples distribution. A minimum discrepancy sequence aims to construct a deterministic sequence of samples to finally produce a uniform space-filling distribution. They are suitable for models that are built incrementally. Typical examples of the methods are Hammersley, Sobol, Halton, and Faure sampling sequences.

### V. Survey over Surrogate-Based Optimization Work

There has been an increasing interest in the topic of incorporating surrogate models with optimization algorithms. Etman [13] applied both quadratic regression and kriging surrogates in some basic structural design problems. He was not satisfied with using kriging approach in design optimization problems. He claimed that it involved impractically expensive computations. However, he highlighted the advantages of flexibility and accurate prediction related to kriging. In another field, Giunta [12] searched for the optimum wing design of a high speed civil transport (HSCT) aircraft for four separate objectives. Quadratic regression and Gaussian kriging surrogates were used. Two test cases were studied involving five and ten design parameters. The design space was constructed using full factorial and D-optimal designs. Again, it was argued that quadratic regression gave better results. Nevertheless, it was stated that more study for kriging was needed to accurately judge it.

However, by time, researchers became less satisfied with the accuracy of polynomial surrogates. It was believed that polynomial regression surrogates are incapable of accurately replacing the exact function over the whole design space. Consequently, many attempts were made to *resize* the design space. Using this concept, the work done by Giunta was improved in [14]. A quadratic polynomial surrogate was used in a multidisciplinary optimization of the HSCT for structural and aerodynamic objectives. In this work, rather than studying the whole design space, only the portion that incorporates the feasible designs. This "reasonable design" approach was proved to yield a considerable improvement in the surrogate accuracy and optimization results. However, we think that such approach can only be efficient in cases where the range of reasonable designs is known or in design improvement applications. On the other hand, in cases where the best design is unknown, this approach may discard promising areas in the design space.

Instead of focusing on a fixed portion of the design space, an iterative approach was used. In [15], a quasi-Newton optimizer was used to search over a quadratic polynomial surrogate for a subsonic flying wing of maximum lift. The design space, constructed using a central composite design (*CCD*), was sequentially resized around the optimum points until no further improvement in the optimum location was achieved. Similarly, [16], the RAE2822 transonic airfoil was optimized for lift-constrained minimum drag. SQP was used to search over a quadratic polynomial surrogate. *LHD* was used to construct the design space which size was continuously reduced around the optimal points. The process was repeated until convergence. The main drawback of such approach was highlighted; a new design set ought to be constructed after every "re-sizing" of the design space. To remedy this drawback, Wang [17] used an inherited *LHD* sampling where after each resizing, the new sample set inherited the samples already existed from previous iterates. The optimum design of an I-beam for minimum vertical deflection under concentrated loading was found faster and more accurate.

Rather than modifying the design space, Papila and Haftka [18] discussed modeling errors associated with polynomial regression surrogates and the various ways of repairing them.

They argued that although regression surrogates smooth the small error noise, outliers (data with large errors) would deteriorate the accuracy of the surrogate. They suggested that the surrogate accuracy would be improved via the iteratively re-weighted least square fitting (IRLS) where low weights were assigned to data with large errors and the whole data were refitted. The process would continue until convergence.

Researchers also recognized another major drawback of using quadratic regression surrogates is that they fail to model response functions with high degree of nonlinearity [19, 16, 20, 21]. Venter et al. [20] used cubic and quartic polynomials instead of the quadratic one. They used gradient-based optimizer to search the minimum weight of a plate with variable thickness under bending and buckling loads. Highly accurate surrogates and better optimization results were obtained. However, the use of higher order polynomials in rather nonlinear problems would invoke instabilities and false local optimum points [6].

There have been many efforts in using other surrogates to model highly nonlinear problems. For example, in [19] both quadratic regression and neural network surrogates were used in predicting the static longitudinal stability of a high speed civil aircraft. A design space of **13** design parameters was constructed using a central composite fractional factorial design including **283** points. It was argued that in predicting problems with unknown, complicated nature, or expanded ranges, quadratic surrogates failed to fit the problem and would be expensive if more design points were added to increase the accuracy. In addition, Fang et al. [21] conducted a mass-constrained optimization of automobiles structure design for maximum crash worthiness (resistance). A composite objective function of three weighted design objectives involving ten parameters was considered and searched using an *SQP* optimization technique. The design space was constructed with orthogonal arrays including **27** design points. Both quadratic regression and radial basis functions surrogates were used and the relative results were compared. It was argued that quadratic regression failed to model the highly nonlinear problems. Hu et al. [22], conducted two similar structure optimization problems. They used the quadratic regression metamodel with the low scale one including two design parameters and used the Gaussian kriging with the highly nonlinear large scale one with seven design parameters. They assessed the superiority of Kriging metamodel in nonlinear problems over the quadratic regression one.

Researchers attempted to enhance the performance of kriging surrogates in many ways. Two approaches were surveyed in the literature. The first approach is to use the functional analysis of variance (*ANOVA*) to specify the *important* design parameters. Considering only the parameters that influence the response function the most would minimize the problem size and eventually improve the surrogate performance.

In [23] a multiobjective genetic algorithm (*MOGA*) was used to construct a Pareto set based on a quadratic polynomial surrogate to optimize a single-stage, a two-stage turbo-pump, and the NASA rotor67 transonic blade. An *LHD* was used to locate **121** sample points resembling up to **14** design parameters in the design space. Jeong et al. [24] coupled Gaussian kriging surrogate with genetic algorithm search technique to optimize the RAE2822 transonic airfoil shape and flap position for maximum lift-to-drag ratio. Ten design parameters were tested via the analysis of variance and only five of them were found of a reasonable influence on the objective function. A design space of five parameters including **50** design points was constructed using the orthogonal arrays. More points were added in regions of poor accuracy using the expected improvement (*EI*) technique.

In another application, *MOGA* was coupled with a Gaussian kriging surrogate in [25] to construct the Pareto set of non-dominated optimum designs of a subsonic engine nacelle. To minimize the problem size, the **40** design parameters were screened first using the analysis of variance (*ANOVA*) techniques to specify the most effective ones that possess higher weights. The kriging surrogate was iteratively improved via augmenting the design set with points obtained from the *GA* optimization and evaluated by the high fidelity flow solver.

Another approach to improve the kriging surrogate performance is to augment the correlation matrix with first and second order derivatives of the correlation function at all points, thus introducing sensitivity information and improving the surrogate structure [3]. Leary et al. [26] conducted an optimization of a gas turbine nacelle inlet. They compared Gaussian kriging, gradient-enhanced Gaussian kriging, and radial basis functions surrogates. They used sequential *LP $\tau$*  sampling to construct the design space and *BFGS* quasi-Newton algorithm to search the optimum. They concluded that kriging surrogates are more accurate compared to *RBF*. They also highlighted the difficulty of constructing the kriging model and evaluating its parameters; further difficulties were shown when gradient data were introduced. The use of gradient-enhanced kriging was given the term *co-kriging* and was studied in more detail recently in [27]. It was claimed that co-kriging would give better model fitness for highly dimensional phenomena and enable using less training points. Two methods of co-kriging, direct and indirect, were addressed. It was concluded generally that minor improvement was achieved compared to the increased cost and complications of building the models.

In [28], the design of a hybrid single element rocket liquid engine injector was optimized for three different objectives. The Pareto set was constructed using *MOGA* based on a quadratic regression surrogate. The design involved four design parameters. Keane [29] searched the optimum NASA transonic airfoil for minimum drag. A Pareto set based on a Gaussian kriging surrogate was constructed using a non-dominated sorting genetic algorithm (*NSGA-II*). The *LP $\tau$*  sampling technique was used to construct the design space.

In [30, 31], a multiobjective optimization was conducted to a three-element airfoil of a civil aircraft wing. Two objectives were involved, maximum lift at landing and near stall condition. A modified version of *MOGA* was used to locate the Pareto set. The initial training set included **30** points and the expected improvement (*EI*) of both objectives was applied to locate additional **60** points in a 6-design parameters *LHD* space. The work was extended in [32] to a three-dimensional wing and the sweep angle was added to the design parameters. By virtue of their advantages, surrogate models were integrated in multidisciplinary optimization schemes. Simpson et al. [6, 7, 8] conducted a multidisciplinary optimization of an aerospike rocket engine nozzle design. Two disciplines, aerodynamics and structure, and three objectives, maximum thrust, minimum weight, and minimum gross lift-off weight, were involved. Both quadratic regression and Gaussian kriging were applied and the relative results were compared. The design space of three design parameters was constructed by orthogonal array sampling and included **25** design points. They pointed out that kriging was more appropriate for highly nonlinear problems and it gave better results.

In [33], a multidisciplinary optimization framework based on surrogate models was implemented in two design problems; an aircraft concept design including **13** design parameters in three disciplines and an autonomous hovercraft design including **11** design parameters. A generalized reduced gradient (*GRG*) search method was used to search over a surrogate model built with artificial neural networks (*ANN*) approach. In [34], another collaborative multidisciplinary framework in the optimization of a tail-less unmanned aerial

vehicle (UAV). Disciplines included aerodynamic performance, structural analysis, and overall performance. They were all based on quadratic regression surrogates with a trust region approach to refine the design space. They concluded that, due to lack of accuracy, quadratic surrogates could be used as a priori investigation and more accurate base functions ought to be used.

Similarly, Won et al. [35], used a quadratic regression model in a design optimization of a high speed stand-off missile (HSSM) within a multidisciplinary optimization. The design space of ten design parameters including up to **129** points was constructed using a *CCD* technique and an *SQP* optimization scheme was used. They highlighted the errors associated with the regression model used. In [36], a multidisciplinary optimization based on a quadratic regression surrogate was conducted. The weight of a cantilevered composite beam under a parabolic distributed load was minimized in two disciplines; a stress-constrained and a deflection-constrained optimization of three design parameters. In the system level, results that satisfied both disciplines and minimized an objective function were investigated. A uniform *LHD* was used to construct the design space which was explored using a sequential quadratic programming (*SQP*) optimizer.

In [37], a kriging-based *MOGA* was used in multidisciplinary optimization of a small jet aircraft to obtain the non-dominated wing designs. Disciplines included aerodynamics and structure, four objective functions, and **109** design parameters. The design space was constructed using *LHD* sampling and new points were added at locations of maximum expected improvement (*EI*). Recently, the weight of a hypersonic vehicle with air-breathing engine was minimized in [38] where a multi-disciplinary optimization based on Gaussian kriging was conducted.

## VI. Conclusions and General Recommendations

The role of optimization in aerodynamic design and manufacturing is increasing over the years making use of the growing development of *CFD* codes and computation capabilities. The main obstacle facing the experts in this field is the high computation costs, memory and time demands associated with real problems. The introduction of much cheaper surrogates tends to remedy these drawbacks. There is a wide variety of surrogate models each has its own structure, accuracy, pros and cons, and area of application.

In this paper, we surveyed the aspects of surrogate-based aerodynamic design optimization in particular. The associate jargon was explained and the related previous efforts were discussed. We can conclude that, the efficient use of surrogates in aerodynamic design optimization problems implies satisfying the following demands:

- A good understanding of the problem in concern is fundamental. The physical nature of the related phenomena and the mathematical behavior should guide you in selecting the appropriate surrogate.
- Choose the appropriate sampling strategy according to the available computation budget: is it better to use all your computational resources to construct the surrogate or leave some to further improvements? Reserve some resources for surrogate assessment.
- Choose the surrogate model that suits the nature of the case investigated. You may need to test more than one model and choose the more suitable.
- Assess the selected surrogate model and find the best ways to increase the accuracy: confining the design space, adding more training points, etc... You need to check the surrogate response at new points with that of the high fidelity model.

- Select the optimization algorithm that is best suited for your case: local/global search, single design/ population, etc...
- Compare the response of design obtained by the surrogate with that of the high fidelity model. How accurate is the predicted value? Are you satisfied with the performance improvement? Do you need to search for better designs?

## Appendix: List of Abbreviations

<i>ADO</i>	aerodynamic design optimization
<i>AMMF</i>	approximation model management frameworks
<i>ANN</i>	artificial neural networks
<i>ANOVA</i>	analysis of variance
<i>BFGS</i>	Broyden-Fletcher-Goldfarg-Shanno, quasi-Newton search algorithm
<i>CFD</i>	computational fluid dynamics
<i>DFP</i>	Davidon-Fletcher-Powell, quasi-Newton search algorithm
<i>DoE</i>	design of experiments
<i>EA</i>	evolutionary algorithm
<i>EI</i>	expected improvement
<i>FCC</i>	face centered cubic
<i>FFD</i>	full factorial design
<i>GA</i>	genetic algorithm
<i>KG</i>	kriging
<i>LHD, LHS</i>	Latin hypercube design, sampling
<i>MAE</i>	maximum absolute error
<i>MARS</i>	multivariate adaptive regression splines
<i>MLE</i>	maximum likelihood estimation
<i>MMF</i>	model management frameworks
<i>MOGA</i>	multiobjective genetic algorithm
<i>MOOP</i>	multi-objective optimization
<i>NSGA</i>	nondominated sorting genetic algorithm
<i>NSGA-II</i>	fast nondominated sorting genetic algorithm
<i>OA</i>	orthogonal arrays
<i>PFD</i>	partial factorial design
<i>PR</i>	polynomial regression
<i>RBF</i>	radial basis functions
<i>RMS</i>	root mean square
<i>RSM</i>	response surface model
<i>SA</i>	simulated annealing
<i>SMF</i>	surrogate management frameworks
<i>SQP</i>	sequential quadratic programming
<i>SVM</i>	support vector machines

## References

- [1] Myers, R. H., and Montgomery, D. C., "Response Surface Methodology: Process and Product Optimization Using Designed Experiments," Wiley, New York, 1995.
- [2] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–435.
- [3] Keane, A. J. and Nair, P. B., "Computational Approaches for Aerospace Design, the Pursuit of Excellence," John Wiley and Sons, Chichester, 2005.
- [4] Konak, A., Coit, D. W., and Smith, A. E., "Multi-objective Optimization Using Genetic Algorithms: A tutorial," *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 992–1007.
- [5] Jeong, S., Murayama, M and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413 – 420.
- [6] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233 -2241.
- [7] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," *AIAA Paper AIAA-98-4755*, 1998.
- [8] Simpson, T. W., "Comparison of Response Surface and Kriging Models in the Multidisciplinary Design of an Aerospike Nozzle," NASA Langley Research Center Hampton, VA, February 1998.
- [9] Jin, R., Chin, W., and Simpson, T. W., "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria," *AIAA Paper AIAA-2000-4801*, 2000.
- [10] Fang, H., M. Rais-Rohani, M., Liu R.-Z., and Horstemeyer, M. F., "A comparative study of metamodeling methods for multiobjective crashworthiness optimization," *Computers and Structures*, Vol. 83, 2005, pp. 2121–2136.
- [11] Simpson, T. W., Peplinski, J. D., Koch, P. N., and Allen, J. K., "Metamodels for Computer-based Engineering Design: Survey and recommendations," *Engineering with Computers*, Vol. 17, 2001, pp. 129- 150.
- [12] Giunta, A. A., "Aircraft Multidisciplinary Design Optimization Using Design of Experiments Theory and Response Surface Modeling Methods", Ph.D. Dissertation, Faculty of Virginia Polytechnic Inst. And State Univ., Blacksburg, VA., 1997.
- [13] Etman, L. F. P., "Design and Analysis of Computer Experiments: the Method of Sacks et al.," *Engineering Mechanics Report WFW 94.098*, Depart of Mechanical Engineering, Eindhoven University of Technology, 1994.
- [14] Vladimir O. Balabanov, V. O., Giunta, A. A., Golovidov, O., Grossman, B., Mason, W. H., Watson, L. T., and Haftkalf, R. T., " Reasonable Design Space Approach to Response Surface Approximation," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 308-315.
- [15] Sevant, N. E., Bloor, M. I. G., and Wilson, M. J., " Aerodynamic Design of a Flying Wing Using Response Surface Methodology," *Journal of Aircraft*, Vol. 37, No. 4, 2000, pp. 562- 569.
- [16] Vavalle, A. and Qin, N., "Iterative Response Surface Based Optimization Scheme for Transonic Airfoil Design," *Journal of Aircraft*, Vol. 44, No. 2, 2007, 365- 376.
- [17] Wang, G. G., "Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points," *Transactions of the ASME, Journal of Mechanical Design*, Vol. 125, 2003, pp. 210-220.
- [18] Papila, M. and Haftka, R. T., "Response Surface Approximations: Noise, Error Repair, and Modeling Errors," *AIAA Journal*, Vol. 38, No. 12, 2000, pp. 2336- 2343.

- [19] Daberkow, D. D. and Dimitri N. Mavris, D. N., "New Approaches to Conceptual and Preliminary Aircraft Design: A Comparative Assessment of a Neural Network Formulation and a Response Surface Methodology," 1998 World Aviation Conference September 28-30, 1998, Anaheim, CA.
- [20] Venter, G., Haftka, R. T., and Starnes, J. H., "Construction of Response Surface Approximations for Design Optimization," *AIAA Journal*, Vol. 36, No. 12, 1998, pp. 2242- 2249.
- [21] Fang, H., Rais-Rohani, M., Liu, Z., and Horstemeyer, M. F., "A comparative study of metamodeling methods for multiobjective crashworthiness optimization," *Computers and Structures*, Vol. 83, 2005, pp. 2121–2136.
- [22] Hu, W., Enying, L., Li, G. Y. and Zhong, Z. H., "Development of Metamodeling Based Optimization System for High Nonlinear Engineering Problems," *Advances in Engineering Software*, Vol. 39, 2008, pp. 629- 645.
- [23] Lian, Y. and Liou, M., "Multiobjective Optimization Using Coupled Response Surface Model and Evolutionary Algorithm," *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1316-1325.
- [24] Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, 413- 420.
- [25] Song, W. and Keane, A. J., "Surrogate-Based Aerodynamic Shape Optimization of a Civil Aircraft Engine Nacelle," *AIAA Journal*, Vol. 45, No. 10, 2007, pp.2565- 2574.
- [26] Leary, S. J., Bhaskar, A., and Keane, A. J., " Global Approximation and Optimization Using Adjoint Computational Fluid Dynamics Codes," *AIAA Journal*, Vol. 42, No. 3, 2004, pp. 631- 641.
- [27] Laurenceau, J. and Sagaut, P., " Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498-507.
- [28] Goel T., Vaidyanathan, R., Haftka, R. T., Shyy, Y., Queipo, N. V., and Tucker, K., "Response surface approximation of Pareto optimal front in multi-objective optimization," *Comput. Methods Appl. Mech. Engrg.*, Vol. 196, 2007, pp. 879– 893.
- [29] Keane, A. J., "Statistical Improvement Criteria for Use in Multiobjective Design Optimization," *AIAA Journal*, Vol. 44, No. 4, 2006, pp. 879- 891.
- [30] Kanazaki, M., Tanaka, K., Jeong, S., and Yamamoto, K., "Multi-objective Aerodynamic Optimization of Elements' Setting for High-lift Airfoil Using Kriging Model," 44th AIAA Aerospace Sciences Meeting and Exhibit, January 2006, Reno, Nevada.
- [31] Kanazaki, M., Tanaka, K., Jeong, S., and Yamamoto, K., "Multi-Objective Aerodynamic Exploration of Elements' Setting for High-Lift Airfoil Using Kriging Model," *Journal of Aircraft*, Vol. 44, No. 3, 2007, pp. 858- 864.
- [32] Kanazaki, M., Imamura, T., Jeong, S., and Yamamoto, K., "High-lift Wing Design in Consideration of Sweep Angle Effect Using Kriging Model," 46th AIAA Aerospace Sciences Meeting and Exhibit 7 - 10 January 2008, Reno, Nevada.
- [33] Batill, S. M., Stelmack, M. A., and Sellar Jr., R. S., "Framework for Multidisciplinary Design Based on Response-Surface Approximations," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 287- 297.
- [34] Sobieski, I. P. and Kroo, I. M., "Collaborative Optimization Using Response Surface Estimation," *AIAA Journal*, Vol. 38, No. 10, 2000, pp. 1931- 1938.
- [35] Won, H., Levine, S., Pfaender, H. and Mavris, D. N., " Using Response Surface Metamodels to Optimize the Aerodynamic Performance of a High Speed Standoff Missile within a Multi-Disciplinary Environment," Aircraft Technology, Integration, and Operations Technical Forum, 1-3 Oct., 2002, L.A., Ca.

- [36] Zadeh, P. M. and Toropov, V. V., "Multi-Fidelity Multidisciplinary Design Optimization Based on Collaborative Optimization Framework," 9<sup>th</sup> AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Sept. 2002, Atlanta, Georgia.
- [37] Kumano, T., Jeong, S., Obayashi, S., Ito, Y., Hatanaka, K., and Morino, H., "Multidisciplinary Design Optimization of Wing Shape for a Small Jet Aircraft Using Kriging Model," 44<sup>th</sup> AIAA Aerospace Sciences Meeting and Exhibit, January 2006, Reno, Nevada.
- [38] Tsuchiya, T., Takenaka, Y., and Taguchi, H., "Multidisciplinary Design Optimization for Hypersonic Experimental Vehicle," *AIAA Journal*, Vol. 45, No. 7, 2007, pp. 1655-1662.