52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference
 19th
4 - 7 April 2011, Denver, Colorado

AIAA 2011-2150

A Numerical Experiment on Allocating Resources Between Design of Experiment Samples and Surrogate-Based Optimization Infills

Patrick H. Reisenthel[+] and Daniel J. Lesieutre[¶]
Nielsen Engineering & Research, Inc., Santa Clara, California 95054

The aerodynamic design of an asymmetric oversized launch vehicle payload fairing subject to stability constraints is used as an example of a derivative-free, expensive black box function to investigate optimal resource allocation when designers are confronted with limited computational budgets.

Nomenclature

CDF	=	cumulative distribution function
CFD	=	computational fluid dynamics
DOE	=	design of experiments
EI	=	expected improvement function
LHS	=	latin hypercube sampling
N_DOE	=	number of DOE function calls
N_iter	=	total number of function calls
N_infills	=	number of infill function calls
OBJV	=	objective function value
p	=	probability
RBF	=	radial basis function

I. Introduction / Motivation

Surrogate-based optimization is frequently a good choice when dealing with expensive function evaluations.[1] When using sequentially updated surrogates (also referred to herein as metamodels), one can neglect the cost of identifying the parameters of the metamodel as well as the cost of searching the surrogate, in comparison to the effort of obtaining a single real data point using either physics-based analysis or experimentation.

The main steps involved in surrogate-based optimization include the initialization of metamodel parameters, the seeding of the initial design space using design of experiments (DOE) methods, metamodel/surrogate identification, global searching of the current surrogate, and the acquisition/evaluation of new data points ("infill" stage) based on the results of that search. A termination criterion and a determination of final solution quality or confidence in the final real optimum are also required. The choice of a metamodel type must consider a number of issues, including assumption requirements, the expected design landscape complexity (if known), and the time spent training or retraining the metamodel. The optimization suitability of a given type of surrogate model is problem-dependent and is affected by factors such as the model's ability to fit complex local behavior, and the basic effort associated with the infill stage. References 2 and 3 provide excellent reviews of surrogate-based optimization which can be used to this effect.

The question addressed in this paper is that of choosing how to best allocate limited computational resources, specifically the choice between acquiring more DOE points *a priori*, leaving fewer resources for optimization-based infills, or limiting the initial DOE investment with the attending greater reliance on the results of the optimization. Allocating a greater percentage of computational resources to the initial DOE has natural advantages; these include *(i)* a better characterization of the global design space, and *(ii)* the fact that this stage, if the data are computational, can take advantage of parallel processing. The downside of this approach is that, by limiting the number of infill points, one may not have sufficient resources to attain either the true optimum, or solutions that meet certain suitability conditions, i.e., in a practical sense, achieve sufficient *improvement*. Designers customarily rely on prior experience and expert knowledge to contend with this dilemma. Indeed, informal surveys of opinions on this topic can yield anywhere from a 20% recommended investment in the initial DOE, to upwards of 80%, with the

[+] Vice President & Chief Scientist, 2700 Augustine Drive, Suite 200, Santa Clara, CA 95054, Senior Member.
[¶] Senior Research Engineer, 2700 Augustine Drive, Suite 200, Santa Clara, CA 95054, Senior Member.

optimization being considered a final solution "refinement." The latter thinking tends to be more prevalent in organizations dealing with problems that lend themselves to *local* optimization. However, with the advent of recent improvements in robust surrogate-based optimization, the practical use of Efficient Global Optimization (EGO) is becoming more of a possibility. In addition, infill points resulting from the application of EGO algorithms populate the design space in ways that are different from, yet may compete with, DOE, thus possibly challenging the conventional notions of what the "ideal ratio" may be between the number of DOE and infill points.

The results presented below constitute such an example. They were collected as a follow-up to a benchmarking study presented at the 6th AIAA Multidisciplinary Design Optimization Specialist Conference (Ref. 4). This paper presents some of the initial results, which are based on a CFD-based optimization problem offering a realistic level of difficulty, yet that is computationally manageable, to the extent that meaningful statistics can be extracted. Details of the aerodynamic design problem being considered, along with the high-fidelity data description, can be found in Reference 4.

II. Methods

Surrogate model types range in complexity from polynomial regression and moving least squares to Kriging and support vector regression. The approach used in this paper is based on radial basis function (RBF) models, either fixed basis RBF or parametric RBF.[4,5] The choice of these models is motivated by their qualities in terms of modeling ability, flexibility, generalization properties, and the significant advantages they offer in terms of performance.[5] The RBF class of surrogates covers a wide range of methods, from simple fixed basis RBF, to fully parametric RBF approaching the complexity of Kriging. Within this class, one can trade generality for performance, depending on which parameters are solved by optimization (parametric RBF) versus which parameters are fixed (simple RBF). The following describes the methods used in the results of Section III.

A. Radial Basis Function Model

In an N-dimensional design space, a surrogate function $F: \mathrm{R}^N \rightarrow \mathrm{R}$ is constructed by satisfying data constraints at P available data points. If this response surface acts as an interpolant, then the function F is required to satisfy the constraints

$$F(\boldsymbol{X_i}) = Y_i, \quad i = 1,\ldots, P \tag{1}$$

where each $\boldsymbol{X_i}$ is an N-dimensional vector of design variables, and Y_i are the corresponding dependent variables. In the case where F represents, instead, a regression model fit to the data, then the response surface is required to minimize in the least squares sense the distance $\| F(\boldsymbol{X_i}) - Y_i \|$, $i = 1,...,M$, $M \geq P$.

In the radial basis function approach, the metamodel F is expanded into basis functions Φ_k which are radially symmetric about their control point, $\boldsymbol{\Gamma_k}$:

$$F(\boldsymbol{X}) = \sum_k c_k \, \Phi_k(\boldsymbol{X}, \boldsymbol{\Gamma_k}, b_k) \qquad \Phi_k = f(\|\boldsymbol{X} - \boldsymbol{\Gamma_k}\|, b_k) \tag{2}$$

where f is a scalar shape function (for example, a Gaussian), b_k is an adjustable scale or stiffness parameter, and $\|.\|$ designates the Euclidean norm. For example, if f is chosen to be a Gaussian:

$$\Phi_k(\boldsymbol{X}, \boldsymbol{\Gamma_k}, b_k) = \exp\left(-\frac{(\boldsymbol{X} - \boldsymbol{\Gamma_k})^{\mathrm{T}}(\boldsymbol{X} - \boldsymbol{\Gamma_k})}{b_k^2}\right) \tag{3}$$

With the additional assumptions that (a) the stiffness b_k is uniform, and (b) the control points are chosen among the available data points, the linear model regression design matrix equation $[\mathbf{A}][\mathbf{c}] = [\mathbf{Y}]$ is given by:

$$\begin{bmatrix} f(\|\boldsymbol{X}_1 - \boldsymbol{X}_1\|) & f(\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|) & \ldots & f(\|\boldsymbol{X}_1 - \boldsymbol{X}_P\|) \\ f(\|\boldsymbol{X}_2 - \boldsymbol{X}_1\|) & f(\|\boldsymbol{X}_2 - \boldsymbol{X}_2\|) & \ldots & f(\|\boldsymbol{X}_2 - \boldsymbol{X}_P\|) \\ \vdots & \vdots & \vdots & \vdots \\ f(\|\boldsymbol{X}_M - \boldsymbol{X}_1\|) & f(\|\boldsymbol{X}_M - \boldsymbol{X}_2\|) & \ldots & f(\|\boldsymbol{X}_M - \boldsymbol{X}_P\|) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_P \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} \tag{4}$$

In the case where uncertainty intervals for the Y_i are available, and provided these intervals correspond to random uncorrelated noise (variance σ_0^2), the variance of the surrogate prediction at point $\boldsymbol{X}$ is given by

$$\hat{\sigma}^2(\boldsymbol{X}) = \sigma_0^2 . \boldsymbol{\Phi}(\boldsymbol{X})^{\mathrm{T}} (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A})^{-1} \boldsymbol{\Phi}(\boldsymbol{X}) \tag{5}$$

where $\boldsymbol{\Phi(X)} = [\Phi_1, \Phi_2, ..., \Phi_P]^{\mathrm{T}}$ (see Ref. 6).

B. Objective Function

The examples discussed in this paper consider the unconstrained minimization of a computationally expensive function $Y(\boldsymbol{X})$ over a simply connected domain D:

$$\begin{aligned} &\text{minimize} \quad Y(\boldsymbol{X}) \\ &\{x_1, x_2, \ldots, x_N\} \\ &\text{subject to} \quad \boldsymbol{X} \in D \end{aligned} \tag{6}$$

A key aspect of the surrogate-based optimization process is the search of the metamodel, since $F(\boldsymbol{X})$ evaluations cost little in comparison to evaluations of the true computational function $Y(\boldsymbol{X})$.

Additionally, in the case where an uncertainty model (e.g., (5)) is available, it is possible to make use of the expected improvement function (EI) traditionally used in Kriging.[7] This function has shown considerable promise[3] in terms of driving the optimization and being able to balance exploitation and exploration efficiently.

Assuming a normally distributed error at point $\boldsymbol{X}$ with standard error $\sigma(\boldsymbol{X})$, and defining $\boldsymbol{X}^*$ as the current best solution, the expectation of improving upon $F(\boldsymbol{X}^*)$ can be expressed analytically[7,3] as

$$\begin{aligned} \mathrm{EI}(\boldsymbol{X}) \;=\; E\left[\max\left(F(\boldsymbol{X}^*)-F(\boldsymbol{X}),0\right)\right] \;=\; &\left(F(\boldsymbol{X}^*)-F(\boldsymbol{X})\right).\boldsymbol{\psi}\left(F(\boldsymbol{X}^*)-F(\boldsymbol{X}),\sigma(\boldsymbol{X})\right) \\ + \quad &\sigma(\boldsymbol{X}).\boldsymbol{\phi}\left(F(\boldsymbol{X}^*)-F(\boldsymbol{X}),\sigma(\boldsymbol{X})\right) \end{aligned} \tag{7}$$

where $\boldsymbol{\phi}$ is the probability density function

$$\boldsymbol{\phi}(y,\sigma) \;=\; \frac{1}{\sqrt{(2\pi)}}\,\exp\left(-\frac{y^2}{2\sigma^2}\right) \tag{8}$$

and $\boldsymbol{\psi}$ is the normal cumulative distribution function,

$$\boldsymbol{\psi}(y,\sigma) \;=\; \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{y}{\sqrt{2}\sigma}\right)\right] \tag{9}$$

Thus, Problem (6) is replaced by the following alternative problem involving the surrogate F and its standard error σ:

$$\begin{aligned} &\text{maximize} \quad OBJV(\boldsymbol{X}) \\ &\{x_1, x_2, \ldots, x_N\} \\ &\text{subject to} \quad \boldsymbol{X} \in D \end{aligned} \tag{10}$$

where the objective function value $OBJV(\boldsymbol{X})$ is either $\{\mathrm{EI}(\boldsymbol{X})\}$ if an uncertainty model $\sigma(\boldsymbol{X})$ is available, or $\{-F(\boldsymbol{X})\}$ otherwise.[‡]

C. Update/Infill Criterion

Problem (6) is solved by making use of a sequentially constructed radial basis function surrogate $F(\boldsymbol{X})$. The surrogate is initialized using design of experiments techniques. The metamodel search algorithm used in this study is a form of generalized pattern search[2] which uses a gradient-free, multistart, steepest ascent, hill climbing algorithm to maximize the expected improvement. At each iteration of the optimization, local optima resulting from the metamodel search are sorted based on objective function performance, and the top m virtual performers ($\boldsymbol{X_i} = \mathtt{argmax}[OBJV]$, $i = 1,\ldots,m$) are then used to spawn m new computational analyses of the real function, i.e., $Y(\boldsymbol{X}_{i=1,\ldots,m})$ is subsequently obtained.

These new computational analyses are used to augment the RBF set of potential regressors $\Phi_{k=1..,P}$ as well as the set of potential constraints $\{\boldsymbol{X}, Y(\boldsymbol{X})\}_{i=1,\ldots,M}$. In addition, the metamodel parameters b_k can be dynamically adjusted based on cross-validation error minimization or other criteria. The update stage therefore results in solving a new linear system, similar to (4), for the new surrogate $F(\boldsymbol{X})$. The results presented in this paper use $m = 1$, corresponding to a standard, serial implementation of what Ref. 3 refers to as a "two-stage" infill.

III. Results

The CFD-based aerodynamic problem considered as a test case for this study was described in detail in Ref. 4. This test problem (the aerodynamic design of an asymmetric launch vehicle fairing) was chosen because of its suitable complexity and relevance to the aerospace community. Highlights and key characteristics of this problem are summarized below:

- 5 independent variables describing the outer mold line of the fairing,
- 2 dependent variables (corresponding to the overall pitching moment at two different Mach numbers / angle of attack combinations and treated as individual subsystems),

[‡] Note that, in the case where $\sigma(\boldsymbol{X})$ is not available, using $OBJV(\boldsymbol{X}) = -F(\boldsymbol{X})$ is equivalent to using $OBJV(\boldsymbol{X}) = \mathrm{EI}(\boldsymbol{X})$ with $\sigma_0 = 0$. Tests using infinitesimally small input variance (e.g., $\sigma_0 = 10^{-10}$) have been shown (Ref. 4) to produce virtually identical results, although using $OBJV(\boldsymbol{X}) = -F(\boldsymbol{X})$ is more computationally efficient.

- nonaxisymmetric shock structure and three-dimensional flow separation and reattachment, resulting in highly nonlinear, highly sensitive responses for each subsystem,
- stability constraint enforced via nonlinear objective function (used to combine the two subsystem response surfaces into one figure of merit, *OBJV*),
- large number of local minima,
- acceleration via response surface fit of 640 Navier-Stokes calculations.

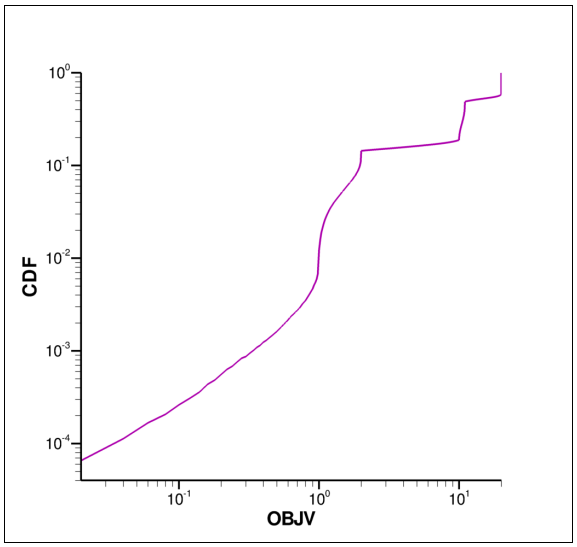

Figure 1. Cumulative distribution function of the objective function *OBJV* based on 1.4 million DOE samples.

In this study, the efficiency of optimization (measured by the number of expensive function evaluations *N_iter*) was considered in the statistical sense, i.e., by repeating each optimization 100 times from different random initial conditions, or, more specifically, by using 100 different realizations of the DOE, for a given number of DOE points. The number of DOE points, *N_DOE*, was either 500, 50, or 15. The DOE realizations thus collected resulted in 1,375,000 samples, the cumulative distribution of which is shown in Fig. 1 which, therefore, characterizes the reference data set.

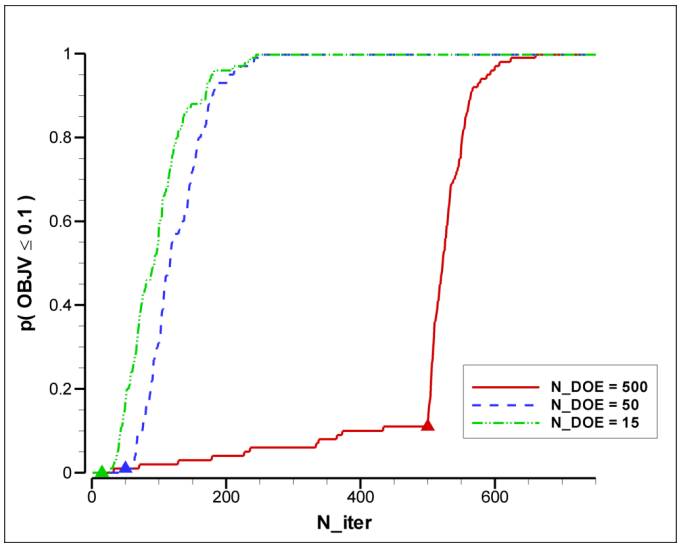

Figure 2. Probability of success (defined as *OBJV* ≤ 0.1) based on 100 independent realizations. The ▲ symbols indicate the beginning of the optimization/infill stage; prior iterations represent DOE samples.

Consistent with Ref. 4, the design optimization is considered successful if the value of the objective function *OBJV at real points* is below a threshold value of 0.5, and is considered highly successful if $OBJV \le 0.1$. The interest, therefore, is in identifying designs which lie in the tail end (lower-left portion) of the distribution shown in Fig. 1. As previously observed,[4] the number of function evaluations required to reach this condition depends not only on the optimization method chosen, but also depends sensitively on the initial design of experiment. It is, therefore, necessary to consider the relative success of one strategy over another in a statistical sense, by means of ensemble processing.

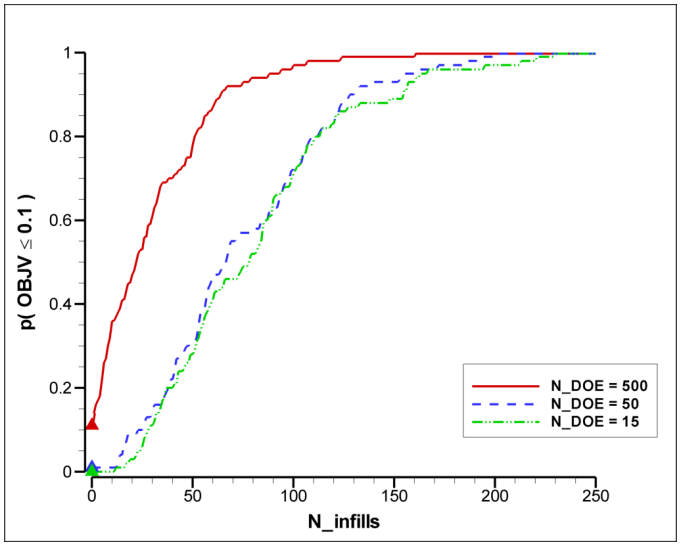

Figure 3. Infill-stage probability of success (defined as *OBJV* ≤ 0.1) based on 100 independent realizations when starting from different numbers of DOE points.

To this end, 100 different realizations of the initial DOE were considered, and, from each of these 100 DOE, a sequential surrogate-based optimization was initiated. The results of these 100 design histories were postprocessed to determine the empirical probability of obtaining real designs reaching the objective (e.g., $OBJV \le 0.1$) as a function of the total number, $N\_iter = N\_DOE + N\_infills$, of expensive function calls. For a given number of function calls, the empirical probability of success is defined as the number of successful optimizations divided by 100 (the total number of trials). Provided one chooses an optimization method that does not get "stuck" in a local optimum, the probability eventually reaches 1.0 after a sufficient number of optimization steps (and, thus, infill points) have been executed.

An example of this empirical probability is given in Fig. 2 for the three values of initial DOE cardinality, *N_DOE*. The change in slope at the onset of the infill stage (marked by a ▲ symbol) is evident, and indicates an acceleration of the rate at which acceptable designs are found as a result of the directed search. It is noteworthy that, for this optimization problem, the *rate* of success is only marginally better for $N\_DOE = 500$, compared to $N\_DOE = 50$ and $N\_DOE = 15$. This is especially visible in Fig. 3, which shows the same data plotted as a function of the number of infill points, *N_infills*.

While one would generally expect an optimal trade-off between the initial cost (*N_DOE*) and subsequent savings due to a greater post-DOE efficiency (higher $dp/dN\_infills$), it is clear (Fig. 2) that this trade-off occurs early (i.e., for a small number of DOE points). The lack of smoothness in the present results, due to an insufficient number of realizations, precludes a precise determination of the optimal *N_DOE*, however.

Similar results, corresponding to the different "success" threshold definitions $OBJV \le 0.2$ and $OBJV \le 0.5$ are shown in Figs. 4 and 5, respectively. In all cases, the observations are similar: increasing the initial DOE investment from 15 or 50 to 500 does not pay off: a given probability of success can be reached 75 to 100 infills sooner at $N\_DOE = 500$, compared to the lower *N_DOE*, but these savings come at the price of a corresponding initial "handicap" of, respectively, 485 and 450 DOE function calls.

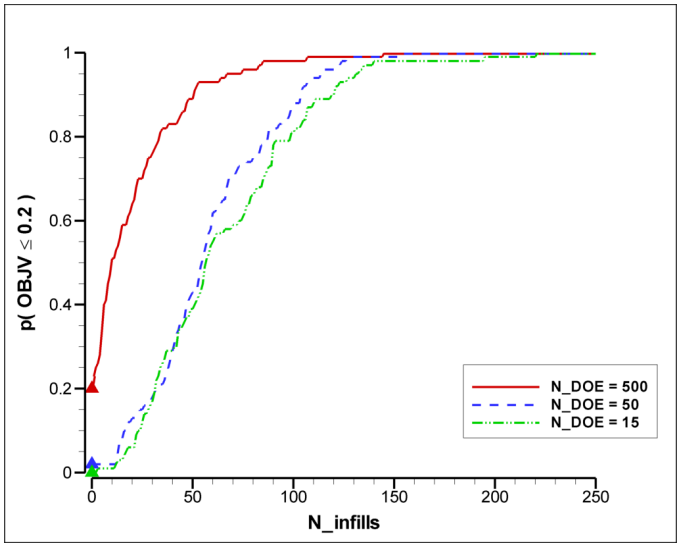

Figure 4. Infill-stage probability of success (defined as *OBJV* ≤ 0.2) based on 100 independent realizations when starting from different numbers of DOE points.

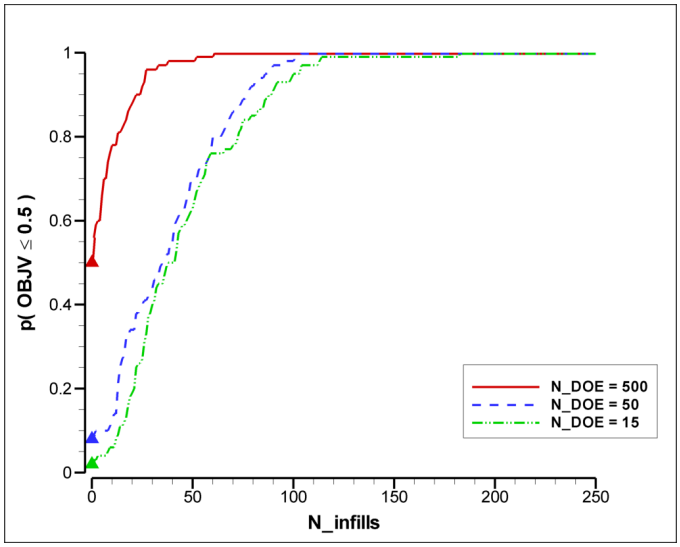

Figure 5. Infill-stage probability of success (defined as *OBJV* ≤ 0.5) based on 100 independent realizations when starting from different numbers of DOE points.

Of course, the above remarks implicitly assume a serial execution of the costly function evaluations, such as might perhaps be the case for large-scale physical experiments. In a high-performance computer setting, the DOE stage can often be performed in parallel and, thus, the conclusions are reversed if clock time is the primary cost metric. Specifically, when starting with a larger *N_DOE*, a higher starting probability of success (the result of a denser sampling of the design space) and the improved slope d*p*/d*N_infills* both combine to create greater efficiency, as shown in Figs. 3-5.

IV. Concluding Remarks / Discussion

The present asymmetric fairing optimization results indicate that, in the absence of parallelism, it is more efficient to devote computational resources to the infill stage of surrogate-based optimization than to its DOE counterpart. This is an intriguing and somewhat counter-intuitive finding which warrants further investigation/confirmation. It is recommended that additional examples drawn from the optimization literature be considered for similar experiments, in order to ascertain the generality of these conclusions.

References

[1]Viana, F., Haftka, R., and Watson, L., "Why Not Run the Efficient Global Optimization Algorithm with Multiple Surrogates?" AIAA-2010-3090.

[2]Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, 2005, pp. 1-28.

[3]Forrester, A. I. J. and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Prog. Aero. Sci.*, Vol. 45, Issues 1-3, Jan.-Apr. 2009, pp. 50-79.

[4]Reisenthel, P.H. and Lesieutre, D.J., "Statistical Benchmarking of Surrogated-Based and Other Optimization Methods Constrained by Fixed Computational Budget," AIAA-2010-3088.

[5]Reisenthel, P. H., Allen, T. T., Lesieutre, D. J., and Lee, S. H., "Development of Multidisciplinary, Multifidelity Analysis, Integration and Optimization of Aerospace Vehicles," NEAR TR 657, Nielsen Engineering & Research, Santa Clara, CA, 2010.

[6]Reisenthel, P. H., Love, J. F., Lesieutre, D. J., and Childs, R. E., "Cumulative Global Metamodels with Uncertainty - a Tool for Aerospace Integration," *The Aeronautical Journal*, Vol. 110, No. 1108, Jun. 2006, pp. 375-384.

[7]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *J. Global Opt.*, Vol. 13, 1998, pp. 455-492.