# BUILDING SURROGATE MODELS BASED ON DETAILED AND APPROXIMATE SIMULATIONS

**Zhiguang Qian*, Carolyn Conner Seepersad**, V. Roshan Joseph *,**

**Janet K. Allen **, and C.F. Jeff Wu *[1]**

*The School of Industrial and Systems Engineering and

**The Systems Realization Laboratory, The George W. Woodruff School of Mechanical Engineering

Georgia Institute of Technology, Atlanta, GA 30332

## ABSTRACT

Preliminary design of a complex system often involves exploring a broad design space. This may require repeated use of computationally expensive simulations. To ease the computational burden, surrogate models are built to provide rapid approximations of more expensive models. However, the surrogate models themselves are often expensive to build because they are based on repeated experiments with computationally expensive simulations. An alternative approach is to replace the detailed simulations with simplified approximate simulations, thereby sacrificing accuracy for reduced computational time. Naturally, surrogate models built from these approximate simulations will also be imprecise. A strategy is needed for improving the precision of surrogate models based on approximate simulations without significantly increasing computational time. In this paper, a new approach is taken to integrate data from approximate and detailed simulations to build a surrogate model that describes the relationship between output and input parameters. Experimental results from approximate simulations form the bulk of the data,

[1] Professor and Corresponding Author
Phone: (404)385-4262; Fax: (404)894-2301; Email: jeffwu@isye.gatech.edu

and they are used to build a model based on a Gaussian process. The fitted model is then 'adjusted' by incorporating a small amount of data from detailed simulations to obtain a more accurate prediction model.

The effectiveness of this approach is demonstrated with a design example involving cellular materials for an electronics cooling application. The emphasis is on the method and not on the results *per se*.

## NOMENCLATURE

| | |
|---|---|
| $D$ | Fixed overall depth of heat exchanger |
| $H$ | Fixed overall height of heat exchanger |
| $k$ | Thermal conductivity of the solid material |
| $\dot{m}$ | Total mass flow rate of cooling fluid |
| $n$ | Number of sample points |
| $n_a$ | Number of sample points for approximate simulations |
| $n_d$ | Number of sample points for detailed simulations |
| $\mathbf{p}$ | Power parameters for a correlation function |
| $\dot{Q}$ | Total rate of steady state heat transfer |
| $\mathbf{R}$ | Correlation matrix |
| $\mathbf{R}_a$ | Correlation matrix for an approximate simulation |
| $\mathbf{R}_\delta$ | Correlation matrix for $\delta(\mathbf{x})$ |
| $R(\mathbf{x}, \mathbf{x}')$ | Correlation between points $\mathbf{x}$ and $\mathbf{x}'$ |
| $T_{in}$ | Inlet temperature of the cooling fluid |
| $T_{wall}$ | Temperature of the heat source |
| $W$ | Fixed overall width of the heat exchanger |
| $x_i$ | Input variable $i$ |
| $y(\mathbf{x})$ | Output of a computer simulation at input value $\mathbf{x}$ |
| $y_a(\mathbf{x})$ | Output of an approximate simulation at input value $\mathbf{x}$ |
| $y_d(\mathbf{x})$ | Output of a detailed simulation at input value $\mathbf{x}$ |

$\boldsymbol{\delta}$ $\mathbf{y}_d - \rho(\mathbf{x})\mathbf{y}_a$

$\boldsymbol{\beta}$ Coefficients of mean function

$\boldsymbol{\beta}_a$ Coefficients of mean function for $y_a(\mathbf{x})$

$\delta_0$ Constant mean function of $\delta(\mathbf{x})$

$\rho(\mathbf{x})$ Scale adjustment term

$\rho_i$ Liner coefficients for $\rho(\mathbf{x})$

$\sigma^2$ Variance of a stationary Gaussian process

$\sigma_a^2$ Variance of $y_a(\mathbf{x})$

$\boldsymbol{\theta}$ Scale correlation parameters

$\boldsymbol{\theta}_a$ Scale correlation parameters of $y_a(\mathbf{x})$

$\boldsymbol{\theta}_\delta$ Scale correlation parameters of $\delta(\mathbf{x})$

**KEYWORDS:** Surrogate models, metamodels, conceptual design

## 1. FRAME OF REFERENCE

Preliminary design of a complex system often involves exploring a broad design space or region of design variable values. Many detailed analysis programs are available for use in the latter stages of design, but they can be extremely expensive for exploring broad regions. One solution has been to simplify the simulations and obtain data from more approximate simulations. For these approximate simulations, accuracy is sacrificed to reduce computational time. However, when it is desirable to explore a large design space that includes broad ranges of design variables, repeated approximate simulations still generate substantial computational loads.

Another approach is to create surrogate models to replace individual simulations. These surrogate models have been used widely in design. Computer experiments in which the design variables cover a carefully chosen range of values are used to create the surrogate models. Values of the design variables are

chosen in specific patterns called experimental designs[1-2] and performance is simulated at these points. The responses and input values are combined statistically to create functional relationships between input variables and performance; these functional relationships are the surrogate models. The surrogate models can be used for robust design[3] or linked to optimization routines, or they can serve as a bridge for integration across multiple functions[4] or across different levels of abstraction[5].

Familiar methods for creating surrogate models include response surface modeling[6] and kriging[7-9] an example of their use in design is presented by Chen and coauthors[3]. However a wide variety of techniques are available[10]. In addition to the choice of the metamodeling method, the accuracy of a surrogate model is determined by the experimental design used to select data points, the size of the design space or range of explored values of design variables, the accuracy of the simulation at each data point and the numbers of data points available to compute the surrogate model[10].

In the last decade, methods for improving the accuracy and computational efficiency of metamodeling procedures have been actively studied. One approach has been to successively reduce the design space, thus simultaneously reducing the extent of the approximation of the metamodels. There are several ways to accomplish this, including the use of trust regions[11-14], heuristics[15], move limits[16] and by using an adaptive response surface method in which the design space is systematically reduced by discarding regions with large objective function values at each modeling-optimization iteration[17-18]. Entropy maximization has also been studied[19-20]. Wang and Simpson propose an intuitive metamodeling method based on hierarchical fuzzy clustering which helps a designer reduce metamodels to regions of interest to a designer[21].

Another way of reducing the design space is by reducing its dimensionality[22]. Typically, the design space is screened to identify and remove design variables which are less important. However, it can be difficult to obtain substantial reductions of dimensionality for large-scale problems[23]. Super-efficient screening methods for removing less important design variables are also available. Both group-screening[24] and sequential bifurcation[25-26] must be applied cautiously for designs in which multiple responses are considered; screening using supersaturated statistical experimental designs is preferable for situations with multiple responses[27-28].

We believe that the choice of metamodeling method must take into consideration both computational time and metamodel accuracy because different aspects of metamodeling may be important in different circumstances. Our method involves creating metamodels based on both approximate and detailed (accurate) simulations and thus using information that is developed necessarily when creating the simulations; a preliminary report of our approach has appeared[29]. Osio and Amon[30-31] also propose a multistage kriging method to sequentially update and improve model accuracy. This method is compared with our approach in greater detail in Section 2.4.

In general there is a trade-off between the accuracy of a surrogate model and the resources needed to build it. If surrogate models are built with a reduced number of data points, they are generally less accurate than models built with a larger number of data points. If detailed, computationally expensive simulations are replaced with approximate simulations, many more data points can be obtained. However, a surrogate model built with approximate information may produce biased results. A practical, alternative strategy is to run a large number of approximate simulations and a smaller number of detailed simulations and then combine the two sets of results to produce a final surrogate model.

In this paper, we develop a framework in which we can combine results from both detailed simulations and approximate simulations to create surrogate models, which are as accurate as possible, given the resources available. Since the approximate simulations form the bulk of the data, they are used to build a model based on a Gaussian process that assumes a simple mean part with a flexible "residual" part. The fitted model is then "adjusted" by incorporating information from the detailed simulations.

In Section 2, we briefly review our approach along with the procedure of Gaussian process modeling that is foundational to it. As an illustration, we apply this approach for designing linear cellular alloys in Section 3. Discussions and possible extensions of our approach are presented in Section 4.

## 2. BUILDING A SURROGATE MODEL BASED ON DETAILED AND APPROXIMATE SIMULATIONS

Integration of results from detailed simulations (DS) and approximate simulations (AS) is not a straightforward task because the two sets of results have significantly different distributional assumptions. One possible way to combine the AS and DS data is to link them by a simple structure and then build a

prediction model for DS directly. This one-step approach has one major disadvantage. Due to the paucity of the DS runs, the resulting surrogate model can be very imprecise and can lead to inaccurate predictions. To overcome this problem and create an accurate surrogate model, we propose a novel *two-step approach* based on Gaussian process modeling. In this work, we assume that the DS produces results that are in agreement with the results from the true process. Thus, we neglect the error in the DS results compared to the true process. This is a reasonable assumption in many computer experiments including the example in Section 3. Thus, the objective is to create a surrogate model that can produce predictions close to the DS results.

A generic diagram is presented for the new two-stage approach in Fig. 1. Stage 1 involves designing and generating computer experiments for detailed and approximate simulations. Key to the approach is Stage 2—a novel two-step modeling strategy. This sets our method apart from existing surrogate model building techniques. The basic idea is to use AS results to provide a *base surrogate model* and adjust the model by DS results. The detailed description of these two steps will be given in Sections 2.3 and 2.4, respectively. Stage 3 consists of the application part of the procedure. When a *final surrogate model* is available, various further investigations, such as optimization, sensitivity analysis, and calibration can be performed here.

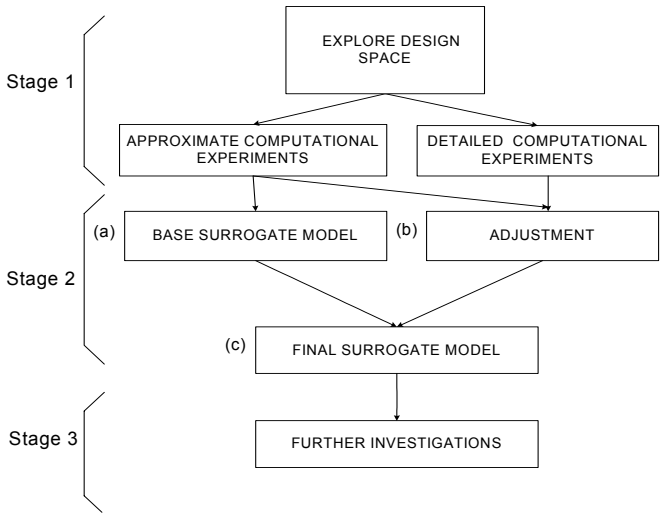


**Fig. 1. Diagram of the proposed approach for combining detailed and approximate data into a surrogate model**

The modeling part of the procedure consists of the following two steps:

(1) Fit a Gaussian process model using only AS data.

(2) Adjust the fitted model in step 1 with DS data.

Since AS results form the bulk of the data, AS results can be used to fit a smooth response surface in the first step. In the second step, this fitted surface is adjusted by DS data, so that the resulting model is close to DS data. The detailed description of these two steps is given in Sections 2.3 and 2.4, respectively.

### 2.1 Gaussian Process Modeling

Gaussian process modeling provides the mathematical foundation for much of the work presented here. Therefore, we start with the rationale for using Gaussian Process modeling and a description of the relevant statistics. The computer experiments considered in this paper are deterministic; there is no random error associated with them, that is, repeated simulations with identical starting points will yield identical results. Due to this deterministic nature, the predicted value from a good surrogate model should equal the actual observed value from simulation, i.e., it should be an *interpolator* to connect all observed data points. This makes modeling computer experiments different from modeling physical experiments, where random noise is present. Traditional modeling methods used to analyze physical experiments, e.g., linear regression and response surface modeling (RSM) often cannot interpolate the observed data unless a model with very high order terms is used. Therefore, they cannot handle the analysis of computer experiments and new methods are needed. In addition to the interpolating property, finding a good surrogate model for complex computer code requires the following: (1) the surrogate model should provide an accurate approximation to the underlying complex pattern of the response, e.g., local increasing and decreasing trends, (2) the computational time required to construct the surrogate model must be low compared to the computational time required to generate data with simulations.

A surrogate model built by Gaussian process modeling satisfies all the above requirements. In addition, this type of model also enables us to provide an error bound for the predicted output at a new input value. Several groups have noticed and used these attractive features[10, 32-35] in various applications.

In a computer experiment, we apply Gaussian process modeling by assuming that the data consist of $n$ vectors of input variable values denoted by $\mathbf{X} = (\mathbf{x}_1^t, \cdots, \mathbf{x}_d^t)^t$ for $d$ covariates and the corresponding response values $\mathbf{y} = (y_1, \cdots, y_n)^t$. The Gaussian process model assumes the following structure:

$$\begin{aligned} y(\mathbf{x}_i) &= \mu_i + \varepsilon(\mathbf{x}_i), \qquad i = 1, \cdots, n, \\ \mu_i &= \sum_{h=0}^{d} \beta_h f_h(\mathbf{x}_i), \end{aligned} \tag{1}$$

where $f_h(\ )$ are pre-specified functions of $\mathbf{x}$, $\beta_h$'s are unknown coefficients and $\varepsilon(\boldsymbol{x})$ is assumed to be a realization of a stationary Gaussian process.

Often a very simple parametric form of $f_h(\cdot)$ is assumed. If $f_h(\mathbf{x}) = x_j$ (and $x_j^2$ respectively), $\beta_h$ represents the linear main effect of $x_j^2$ (and quadratic main effect of $x_j$ respectively).

The specification of $\varepsilon(\mathbf{x})$ is critical here. Unlike in linear regression, the $\varepsilon_i$ 's in Eq. (1) are assumed to be correlated. Any lack of fit in the mean part will be absorbed into $\varepsilon(\mathbf{x})$ as "residual". This "residual" term can be very non-linear and complex indeed. Moreover, as two points $\mathbf{x}_i$ and $\mathbf{x}_j$ get closer together, $\varepsilon(\mathbf{x}_i)$ and $\varepsilon(\mathbf{x}_j)$ must also become closer. Thus, the $\varepsilon_i$ 's must be positively correlated. In addition, if $y(\mathbf{x})$ is continuous in $\mathbf{x}$, so is $\varepsilon(\mathbf{x})$.

The dependence of $\varepsilon(\cdot)$ over different $\mathbf{x}$ values is determined by specifying the following correlation structure:

$$\begin{aligned} \text{cov}(\varepsilon(\mathbf{x}_i, \mathbf{x}_j)) &= \sigma^2 R(\mathbf{x}_i, \mathbf{x}_j), \\ R(\mathbf{x}_i, \mathbf{x}_j) &= \exp[-d(\mathbf{x}_i, \mathbf{x}_j)]. \end{aligned} \tag{2}$$

The correlation function $R(\mathbf{x}_i, \mathbf{x}_j)$ in Eq. (2) is a function of the "distance" between $\mathbf{x}_i$ and $\mathbf{x}_j$. If the "distance" is measured as a Euclidean distance, there will be a tendency to give the same weight to all

variables and therefore the Euclidean distance cannot be used to distinguish different factor effects. To overcome this, the following flexible "weighted" distance function is adopted:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{h=1}^{d} \theta_h \mid \mathbf{x}_{ih} - \mathbf{x}_{jh} \mid^{p_h}, \tag{3}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)$ and $\mathbf{p} = (p_1, \cdots, p_d)$ in Eq. (3) are scale and power parameters, respectively. There are several possible choices for the correlation functions[33] but the structure of Eq. (3) allows each input dimension to have its own scale and power parameters. Varying their values will change the relative importance of factor effects on the response and this specification will also simplify the calculation of model parameters. The Gaussian correlation is for the case $p_h = 2$ and its associated processes are infinitely differentiable in the mean square sense[33]. If $p_h = 1$, the associated processes become the Ornstein-Uhlenbeck process. Typically, $p_h$ is restricted in the interval [1, 2]. Moreover, in this interval, as $p_h$ becomes larger, the sample paths of processes become smoother. The Gaussian correlation fixes $p_h$ at 2 and this simplification reduces the complication of calculating estimates for correlation parameters. As a result, the Gaussian correlation is often adopted in the modeling[34-35]. In the example given in Section 3, we will follow this convention.

The unknown parameters that need to be estimated for the model are $\boldsymbol{\theta}$, $\boldsymbol{p}$ and $\sigma^2$. The maximum likelihood method has been adopted in many cases[34-35]. The log-likelihood, up to an additive constant, is:

$$-\frac{1}{2}[n \ln \sigma^2 + \ln \mid \mathbf{R} \mid - \frac{(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\beta})}{2\sigma^2}], \tag{4}$$

where $\mathbf{R}$ is the $(n \times n)$ matrix with entries $R(\mathbf{x}_i, \mathbf{x}_j)$ for *i, j* = 1,…,*n*, which depend on $\boldsymbol{\theta}$ and $\mathbf{p}$, $\mathbf{F}$ is the $n \times (d+1)$ matrix of regressors having $(i,j)^{th}$ element $f_{(j-1)}(\mathbf{x}_i)$ for $1 \le i \le n, \quad 1 \le j \le (d+1)$, and $\boldsymbol{\beta}$ is a $(d+1) \times 1$ vector of unknown regression coefficients.

Given the values of $\boldsymbol{\theta}$ and $\mathbf{p}$, the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ and $\sigma^2$ can be computed easily:

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^t\ \mathbf{R}^{-1}\ \mathbf{F})^{-1}\ \mathbf{F}^t\ \mathbf{R}^{-1}\ \mathbf{y}\ , \tag{5}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}})^t\ \mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}})}{n} \quad . \tag{6}$$

Substituting $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ into Eq. (4), we obtain a simplified form:

$$-\frac{1}{2}(n \ln(\hat{\sigma}^2) + \ln |\ \mathbf{R}\ |) \quad , \tag{7}$$

where $\hat{\sigma}^2$ and **R** are both functions of $\boldsymbol{\theta}$, **p** and the data. The maximization of Eq. (7) is an optimization problem in the space of $\boldsymbol{\theta}$ and **p** and there may possibly be multiple local maxima. In the example in Section 3, we use the *optim* function[36] to estimate $\boldsymbol{\theta}$ (with **p** set at 2). The *optim* function is based on a limited memory algorithm for bounded constrained optimization[37]. We verified empirically that this *optim* function produces almost the same results as the powerful optimization function *fmincon* in Matlab using the Sequential Quadratic Programming (SQP) algorithm.

The best linear unbiased predictor (BLUP)[33] is adopted to predict $y(\mathbf{x}^*)$ at an untried $\mathbf{x}^*$,

$$\hat{y}(\mathbf{x}^*) = \mathbf{f}_*^{\,t}\hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}\hat{\mathbf{R}}^{-1}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}}), \tag{8}$$

where $\hat{\mathbf{r}}$ is the vector of correlations between $\varepsilon(\mathbf{x}^*)$ and $(\varepsilon(\mathbf{x}_1),\cdots,\varepsilon(\mathbf{x}_n))$, and $\mathbf{f}_* = f(\mathbf{x}^*)$ is the (d + 1)x1 vector of regressors at $\mathbf{x}^*$. Moreover, it can be shown that $\hat{y}(\mathbf{x}_i)$ equals $y_i$. Thus, the BLUP smoothly interpolates all the observed data points. In addition to this interpolating property, it has been well demonstrated that the BLUP can approximate fairly complex functions[33].

### 2.2 Modeling the Approximate Simulation Data

Using the Gaussian process modeling described in Section 2.1, we now develop an approach for building surrogate model. We first build a surrogate model based on the approximate simulations only.

This model will be further refined later. Usually only a constant term (i.e. $\mu_i = \beta_0$ in Eq. (1)) is used in the main part of the Gaussian process model[35]. However, in some circumstances it is reasonable to assume that the factors considered in the experiment have linear effects on the output[38-39]. By following this convention, we choose the model below for the output of the approximate simulation $y_a$ ,

$$y_a(\mathbf{x}) = \beta_{a0} + \sum_{h=1}^{d} \beta_{ah} x_h + \varepsilon_a(\mathbf{x}) \tag{9}$$

where $\beta_{a0} + \sum_{h=1}^{d} \beta_{ah} x_h$ is the linear mean part and $\varepsilon_a(\mathbf{x})$ is the residual part that is assumed to be a stationary Gaussian process with mean zero, variance $\sigma_a^2$ and correlation parameters $\boldsymbol{\theta}_a$. Because a large number of AS runs are available, $(\boldsymbol{\beta}_a, \boldsymbol{\theta}_a, \sigma_a^2)$ can usually be estimated accurately. The BLUP for $y_a(\mathbf{x}^*)$ at an untried $\mathbf{x}^*$ is

$$\hat{y}_a(\mathbf{x}^*) = \mathbf{f}_a^t \hat{\boldsymbol{\beta}}_a + \hat{\mathbf{r}}_a \hat{\mathbf{R}}_a^{-1} (\mathbf{y}_a - \mathbf{F}_a \hat{\boldsymbol{\beta}}_a), \tag{10}$$

where $\mathbf{f}_a$, $\hat{\mathbf{r}}_a$, $\hat{\mathbf{R}}_a$ and $\hat{\mathbf{F}}_a$ are defined as in Section 2.2. Throughout the remaining part of this paper, we shall refer to the model in Eq. (10) as the *base surrogate model*.

### 2.3 Adjustment Based on Detailed Simulation Data

Because AS and DS are generated based on distinct assumptions, their values can be quite different. This is the case in the example analyzed in Section 3. The same input values are used for some of the AS and DS, these data can be used to adjust the base surrogate model. Suppose $n_d$ AS runs share the same input values as $n_d$ DS runs, a very simple adjustment can be done by using a location-scale adjustment, i.e.,

$$y_d(\mathbf{x}_i) = \rho y_a(\mathbf{x}_i) + \delta, i = 1, \cdots, n_d. \tag{11}$$

However, some cases may also exhibit a non-linear discrepancy between AS and DS. As an extension of the above procedure, a more sophisticated adjustment can be obtained by making the following two

changes in Eq. (11): (a) substitute the constant $\rho$ with a linear regression function $\rho(\mathbf{x})$, (b) replace the constant $\delta$ by a Gaussian process $\delta(\mathbf{x})$. These modifications lead to the following model:

$$y_d(\mathbf{x}_i) = \rho(\mathbf{x}_i) y_a(\mathbf{x}_i) + \delta(\mathbf{x}_i), \quad i = 1, \cdots, n_d, \tag{12}$$

where

$$\rho(\mathbf{x}_i) = \rho_0 + \sum_{j=1}^{d} \rho_j x_{ij} \tag{13}$$

is the linear regression function and $\delta(\mathbf{x})$ is assumed to be a stationary Gaussian process with mean $\delta_0$, variance $\sigma_\delta^2$ and correlation parameters $\boldsymbol{\theta}_\delta$.

An assumption is made here to simplify the modeling and the adjustment procedure. We will treat $y_a(\mathbf{x}_i)$ value from step 1 as *fixed*. Thus, with the Gaussian process assumption of $\delta(\mathbf{x})$, the distribution of $\mathbf{y}_d = (y_d(\mathbf{x}_1), \cdots, y_d(\mathbf{x}_{n_d}))^t$ is normal and the log likelihood of $\mathbf{y}_d$, up to an additive constant, can be written as

$$-\frac{1}{2}[n_d \ln \sigma_\delta^2 + \ln |\mathbf{R}_\delta| - \frac{(\mathbf{y}_d - \mathbf{F}_d \boldsymbol{\alpha})^t \mathbf{R}_\delta^{-1} (\mathbf{y}_d - \mathbf{F}_d \boldsymbol{\alpha})}{2\sigma_\delta^2}], \tag{14}$$

where $\mathbf{F}_d$ is the regression matrix $\begin{pmatrix} 1 \ , y_a(\mathbf{x}_1), y_a(\mathbf{x}_1)x_{11}, \cdots, y_a(\mathbf{x}_1)x_{1d} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 1 \ , y_a(\mathbf{x}_{n_d}), y_a(\mathbf{x}_{n_d})x_{n1}, \cdots, y_a(\mathbf{x}_{n_d})x_{n_d d} \end{pmatrix}$

and $\alpha = \left(\delta_0, \rho_0, \rho_1, \cdots, \rho_d\right)^t$ is the collection of unknown parameters associated with the mean part in Eq. (12). The estimates $\hat{\alpha}$ and $\hat{\boldsymbol{\theta}}_\delta$ can be obtained by maximizing the function in Eq. (14). The optimization procedure is very similar to the one described in Section 2.1, so its details are omitted.

For given values of $\hat{\rho}_i$'s $(i = 0, \cdots, d)$, we can compute the values of $\boldsymbol{\delta} = (\delta(\mathbf{x}_1), \cdots, \delta(\mathbf{x}_{n_d}))$ by using

$$\delta(\mathbf{x}_i) = y_d(\mathbf{x}_i) - \hat{\rho}(\mathbf{x}_i) y_a(\mathbf{x}_i), \quad i = 1, \cdots, n_d, \tag{15}$$

where

$$\hat{\rho}(\mathbf{x}^*) = \hat{\rho}_0 + \sum_{j=1}^{d} \hat{\rho}_j x^*_j \tag{16}$$

is the fitted regression function for the scale adjustment.

At an untried point $\mathbf{x}^*$, a BLUP predictor can be constructed as

$$\hat{\delta}(\mathbf{x}^*) = \hat{\delta}_0 + \hat{\mathbf{r}}_\delta \hat{\mathbf{R}}_\delta^{-1} (\boldsymbol{\delta} - \mathbf{F}_\delta \hat{\delta}_0), \tag{17}$$

where $\hat{\mathbf{r}}_\delta$ and $\hat{\mathbf{R}}_\delta$ are defined in Section 2.2, and $\hat{\delta}_0$ is obtained previously as part of $\hat{\alpha}$. The predictor $\hat{\delta}(\mathbf{x}^*)$ in Eq. (17) is used as a building block to establish the final surrogate model.

### 2.4 Building and Evaluating the Final Surrogate Model

Based on the base surrogate model in Eq. (10) and the adjustments results in Eqs. (16) and (17), a simple *plug-in* method is used to establish the *final surrogate model* for an untried $\mathbf{x}^*$,

$$\hat{y}_d(\mathbf{x}^*) = \hat{\rho}(\mathbf{x}^*)\hat{y}_a(\mathbf{x}^*) + \hat{\delta}(\mathbf{x}^*), \tag{18}$$

where $\hat{\rho}(\mathbf{x}^*)$ is the fitted scale adjustment term in Eq. (16), $\hat{y}_a(\mathbf{x}^*)$ is the predicted value from the base surrogate model in Eq. (10), and $\hat{\delta}(\mathbf{x}^*)$ is the fitted location adjustment term in Eq. (17). As mentioned in Section 2.1, the prediction based on the base surrogate model is not very accurate. Because we have adjusted this model using detailed simulation data, the prediction from Eq. (18) be close to the detailed simulations in general. Moreover, the final surrogate model, $\hat{y}_d(\cdot)$ in Eq. (18) smoothly interpolates all the observed the detailed simulation data; this is desirable.

Osio and Amon[30] and Pacheco, Amon and Finger[31] have proposed another Bayesian approach to integrate detailed and approximate simulation data for engineering design. Their methodology consists of two steps: (i) a Gaussian process is used to specify the prior knowledge about the approximate simulation, from which a posterior process about the approximate simulation is obtained (using a standard Bayesian argument); (ii) use the posterior process obtained in (i) as the prior knowledge for the detailed simulation to

obtain a posterior process for the detailed simulation. The mean of the posterior process obtained in (ii) is used as a surrogate model for the detailed simulation. The limitations of this methodology and differences from ours are highlighted below. First, their Bayesian updating starts with the same distribution. Thus the posterior mean for the current stage and the posterior mean for the previous stage have the *same* expectation (see Eq. (11) of Pacheco et al.[31]). This does not reflect the common observation that the detailed simulation data can be closer to the true values than the approximate simulation data. To accommodate this, our approach provides a location-scale adjustment of the approximate simulation values in Eqs. (11)-(13). A pictorial illustration of the adjustment as applied to real data will be presented in the next Section. Second, their approach always produces a smaller posterior variance for the detailed simulations than for the approximate simulations. While this may be observed in some situations, it has no general physical or mathematical justification. Because both approximate and detailed simulations are deterministic, their uncertainties (as rendered by the Bayesian priors) can be more realistically modeled in terms of their distances from the true values. This is why we propose the use of the location-scale adjustment. Finally, their approach requires the specification of a prior for the mean of the approximate. How can such a prior be chosen realistically? The use of a fully Bayesian approach (as described by Currin and coauthors[40]) can solve this problem but the computational price will be prohibitive for engineering problems. To be fair, our approach also needs to specify the values of some process parameters, e.g., the choice of the correlation structure in Eq. (2) with $p_h$ =2. However, our procedure only requires using the maximum likelihood method to estimate k scale parameters for a Gaussian process. This can be computed quickly by using efficient optimization tools such as the *optim* function in R[36].

To illustrate our approach, in the next section we consider the design of a linear cellular material, which is used to dissipate heat from a microprocessor.

## 3. DESIGNING LINEAR CELLULAR MATERIALS WITH THE SURROGATE MODEL BUILDING APPROACH

Consider the design of a heat exchanger for a representative electronic cooling application. As illustrated in Fig. 2, the device is used to dissipate heat generated by a heat source such as a

microprocessor. The mechanism for heat dissipation is forced convection via air with entry temperature, $T_{in}$, in degrees Kelvin and total mass flow rate, $\dot{m}$, measured in kilograms per second. Steady state, incompressible laminar flow is assumed. The device is assumed to have fixed overall width ($W$), depth ($D$), and height ($H$) of 9, 25, and 17.4 millimeters, respectively. It is insulated on the left, right, and bottom sides and is subjected to a heat source at constant temperature, $T_s$, in degrees Kelvin on the top face.

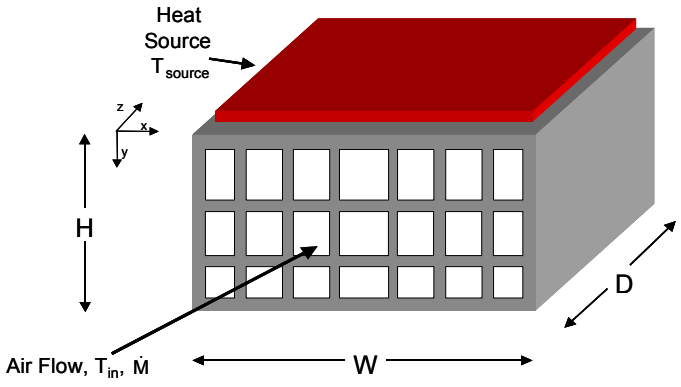


**Fig. 2 Compact, forced convection heat exchanger with graded rectangular linear cellular alloys**

The device is comprised of linear cellular material—ordered, metallic cellular material with extended prismatic cells. These materials can be produced with nearly arbitrary two-dimensional topologies, metallic base materials, and wall thicknesses as small as 50 microns via a thermo-chemical extrusion fabrication process developed at Georgia Tech[41]. Prismatic cellular materials have a combination of properties that make them especially suitable for many multifunctional applications, including actively cooled, lightweight structures[4, 42-44]. Although cell topology and dimensions can be varied, the prismatic cellular material is composed exclusively of rectangular cells for this example. There are four columns of cells with interior cell widths of 2 mm, and three rows of cells with interior cell heights of 10, 5, and 2 mm for the uppermost, middle, and lower rows of cells, respectively. The solid material in the walls of the prismatic cellular material is assumed to have thermal conductivity, $k$, in Watts per meter-Kelvin.

The design objective is to maximize the total rate of steady state heat transfer achieved by the device. Some of the factors affecting this objective include the topology and dimensions of the cells and cell walls,

the flow rate and temperature of the incoming air, the temperature of the heat source, and the thermal conductivity of the solid material in the walls of the device. In other design activities, we have adjusted the dimensions of the device[4]; here, we intend to explore the heat transfer rate as a function of the mass flow rate of entry air, $\dot{m}$, the temperature of entry air, $T_{in}$, the temperature of the heat source, $T_{wall}$, and the solid material thermal conductivity, $k$.

To analyze the impact of these factors on heat transfer rates, we use two types of simulations—computationally expensive FLUENT finite element simulations and relatively fast but more approximate finite difference simulations. Details of the two approaches are available in the literature, but it is important to highlight the costs and benefits of the two approaches in terms of accuracy and computational time. FLUENT is a commercial software package for analyzing fluid flow and heat transfer problems with a computational fluid dynamics (CFD) solver[45]. The finite difference approach is an approximate numerical technique for solving two- or three-dimensional heat transfer problems[46]. Both the finite difference method and FLUENT simulations have been used to simulate the thermal behavior of prismatic cellular materials[4,47]. For examples similar to the present one, each FLUENT simulation requires two to three orders of magnitude more computing time than the corresponding finite difference simulation. For example, the first data point in Table 2 requires 1.75 hours and approximately 2 seconds of computing time for the FLUENT and finite difference simulations, respectively, on a 2.0 GHz Pentium 4 PC with 1 GB of RAM. However, the FLUENT simulations are more accurate than the finite difference simulations by 10 to 15% or more.

Our objective is to build a surrogate model that can be used in the design process and represents the functional relationship between design factors and the total rate of steady state heat transfer. To build the surrogate model, we utilize results from both FLUENT and finite difference simulations. A large number of data points are generated using the finite difference simulation with fewer data points obtained from the FLUENT simulation. We show that even a limited amount of data from FLUENT simulations can be used to improve the accuracy of surrogate models based on approximate finite difference models alone. This is demonstrated in the following subsections.

### 3.2 Generating Design Points for Detailed and Approximated Simulations

An orthogonal array-based Latin Hypercube design[33] with a run size of 64 data points is used to determine the appropriate set of approximate (finite difference) simulations. The assumed ranges of design variables are shown in Table 1. The Latin Hypercube design has good space-filling properties. This can be seen in Fig. 3 in which the four-variable design is projected onto spaces of two variables. For each pair of variables the data points are uniformly distributed in each of the 64 reference square bins. Also, if we divided each bin in Fig. 3 into 8 equally spaced new bins with smaller size (64 new bins in each dimension), we find that each individual variable in each dimension has a nearly uniform distribution in these 64 bins. Among these 64 approximate simulation experiments, results for detailed simulations are generated for 22 of them. Sixteen of the twenty-two experiments are identified using a simulated annealing algorithm and a minimax distance criterion[33]. The remaining six detailed simulation experiments were chosen with a roughly uniform distribution in the portion of the design space in which the value of air flow rate, $\dot{m}$, of entry air is small. Background information suggests that there may be a special relationship between the detailed (FLUENT) results, $y_d$, and the approximate (finite difference) results, $y_a$, in this subregion.. The six additional points are added to explore this relationship. The sample data and corresponding response values are listed in Table 2. In this table, the results for the 64 approximate experiments are shown in the $y_a$ column, and the 22 detailed simulation experiments are listed in the $y_d$ column. It is clear from Table 1 that the four input variables have very different scales. These variables are standardized (subtracting their means and multiplying by the reciprocal of their standard deviations) before the analysis.

**Table 1. Assumed ranges for design variables values**

| | Design Variables | | | |
|---|---|---|---|---|
| | $\dot{m}$ (kg/s) | $T_{in}$ (K) | $k$ (W/mK) | $T_{wall}$ (K) |
| Lower Bound | 0.00055 | 270.00 | 202.4 | 330 |
| Upper Bound | 0.001 | 303.15 | 360.0 | 400 |

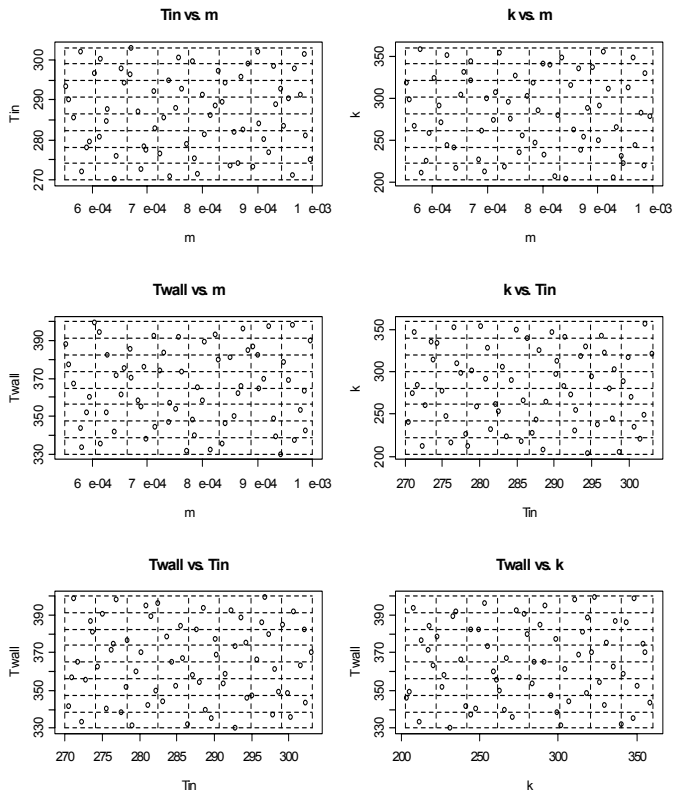


**Fig. 3. 64 points of an orthogonal array-based Latin Hypercube sample. In each plot, there is one point in each of the square bins bounded by dashed lines.**

**Table 2. Sample data for approximate and detailed simulations**

| | Design variables | | | | Responses | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Run | $\dot{m}$ (kg/s) | $T_{in}$ (K) | $k$ (W/mK) | $T_{wall}$ (K) | $y_a$ | $y_d$ |
| 1 | 0.000552 | 293.53 | 318.63 | 388.29 | 25.61 | 23.54 |
| 2 | 0.000557 | 290.18 | 298.27 | 377.49 | 23.24 | |
| 3 | 0.000566 | 285.77 | 266.71 | 367.27 | 21.23 | 20.15 |
| 4 | 0.000578 | 302.17 | 358.13 | 343.72 | 11.44 | 10.17 |
| 5 | 0.000580 | 272.26 | 211.71 | 333.65 | 15.03 | 15.29 |
| 6 | 0.000589 | 278.16 | 225.78 | 351.83 | 18.55 | 18.39 |
| 7 | 0.000594 | 279.54 | 258.51 | 360.13 | 20.74 | 20.52 |
| 8 | 0.000603 | 296.75 | 323.15 | 399.45 | 28.40 | |
| 9 | 0.000612 | 280.83 | 291.53 | 394.72 | 30.22 | 30.12 |
| 10 | 0.000615 | 300.28 | 270.74 | 335.79 | 9.53 | |
| 11 | 0.000626 | 284.89 | 350.46 | 352.29 | 18.13 | 18.17 |
| 12 | 0.000627 | 287.60 | 243.96 | 382.54 | 25.02 | 24.68 |
| 13 | 0.000639 | 270.45 | 241.21 | 341.81 | 17.92 | 19.05 |
| 14 | 0.000643 | 276.17 | 216.99 | 371.60 | 24.20 | 24.96 |
| 15 | 0.000652 | 298.04 | 303.96 | 361.58 | 17.47 | 16.95 |
| 16 | 0.000657 | 294.24 | 330.63 | 375.53 | 22.48 | 22.3 |
| 17 | 0.000669 | 296.33 | 343.16 | 385.81 | 25.07 | |
| 18 | 0.000670 | 303.07 | 321.41 | 370.48 | 18.93 | |
| 19 | 0.000683 | 287.05 | 227.31 | 358.24 | 18.61 | |
| 20 | 0.000689 | 272.70 | 260.91 | 355.37 | 21.31 | |
| 21 | 0.000694 | 278.35 | 212.79 | 376.24 | 25.11 | |
| 22 | 0.000698 | 277.52 | 299.39 | 338.40 | 16.02 | |
| 23 | 0.000711 | 292.26 | 273.31 | 392.54 | 27.47 | |
| 24 | 0.000714 | 283.08 | 306.69 | 344.34 | 16.43 | |
| 25 | 0.000722 | 276.53 | 353.75 | 374.41 | 26.50 | |
| 26 | 0.000730 | 285.51 | 217.74 | 383.92 | 25.88 | |
| 27 | 0.000738 | 295.01 | 295.02 | 347.22 | 14.37 | |
| 28 | 0.000741 | 270.95 | 275.19 | 356.87 | 22.36 | |
| 29 | 0.000751 | 287.99 | 326.02 | 354.08 | 18.17 | 19.57 |
| 30 | 0.000757 | 300.64 | 235.03 | 391.68 | 14.37 | |
| 31 | 0.000763 | 292.82 | 254.84 | 373.38 | 21.96 | 23.33 |
| 32 | 0.000772 | 278.93 | 301.75 | 331.55 | 14.02 | |
| 33 | 0.000782 | 299.86 | 317.84 | 348.41 | 13.68 | |
| 34 | 0.000786 | 275.51 | 247.29 | 340.19 | 16.82 | |
| 35 | 0.000791 | 271.64 | 284.88 | 365.09 | 25.06 | |
| 36 | 0.000800 | 291.42 | 341.48 | 358.59 | 18.83 | |
| 37 | 0.000803 | 281.47 | 232.64 | 389.46 | 28.69 | |
| 38 | 0.000814 | 286.39 | 339.92 | 332.40 | 12.68 | 14.36 |
| 39 | 0.000823 | 288.53 | 207.55 | 393.49 | 27.96 | |
| 40 | 0.000828 | 297.33 | 280.13 | 379.86 | 23.17 | |
| 41 | 0.000836 | 289.62 | 347.65 | 335.44 | 12.79 | |
| 42 | 0.000842 | 294.39 | 203.45 | 346.05 | 13.75 | 15.12 |
| 43 | 0.000851 | 273.71 | 315.27 | 381.14 | 29.08 | 34.8 |
| 44 | 0.000857 | 282.12 | 262.30 | 350.10 | 18.25 | 21.31 |
| 45 | 0.000865 | 274.35 | 335.16 | 362.30 | 23.89 | |
| 46 | 0.000870 | 295.76 | 237.65 | 366.25 | 19.36 | |
| 47 | 0.000874 | 282.50 | 253.25 | 396.36 | 30.90 | 36.11 |
| 48 | 0.000882 | 299.22 | 288.45 | 385.07 | 24.45 | 27.36 |
| 49 | 0.000891 | 273.43 | 336.04 | 386.95 | 31.05 | |
| 50 | 0.000901 | 302.02 | 249.57 | 382.33 | 22.64 | |
| 51 | 0.000903 | 284.25 | 290.90 | 364.99 | 22.22 | 25.37 |
| 52 | 0.000911 | 280.17 | 355.34 | 370.03 | 25.03 | |
| 53 | 0.000920 | 276.89 | 310.73 | 397.78 | 33.27 | |
| 54 | 0.000929 | 298.65 | 205.40 | 349.02 | 13.67 | |
| 55 | 0.000934 | 288.86 | 265.53 | 339.54 | 13.89 | |
| 56 | 0.000943 | 292.77 | 231.01 | 330.19 | 10.16 | |
| 57 | 0.000947 | 283.62 | 222.95 | 378.66 | 25.48 | |
| 58 | 0.000956 | 290.33 | 312.97 | 368.96 | 22.22 | |
| 59 | 0.000964 | 271.23 | 348.00 | 398.52 | 35.05 | |
| 60 | 0.000968 | 297.80 | 244.50 | 337.41 | 10.99 | |
| 61 | 0.000979 | 291.21 | 283.10 | 353.60 | 17.45 | |
| 62 | 0.000985 | 301.50 | 220.37 | 363.20 | 17.14 | |
| 63 | 0.000987 | 281.11 | 329.45 | 342.32 | 16.95 | |
| 64 | 0.000996 | 275.01 | 278.27 | 390.35 | 31.35 | |

### 3.3 Building a Base Surrogate Model

The first step is to build a surrogate model using the approximate simulation results only. Based on background knowledge of the physics of this problem, we know that a linear relationship may exist in this case between the response and the four factors. As a result, a linear structure is included when modeling the mean part of the Gaussian process in Eq. (9). As described in Section 2, the maximum likelihood method is used for estimation. Because $\hat{\sigma}_a^2 = 3.352$ is quite small, we have a good fit. Table 3 lists the linear main effects $\hat{\beta}_{ai}$ for $i = 1,\ldots,4$ (corresponding to $\dot{m}$, $T_{in}$, $k$, and $T_{wall}$, respectively) and the p-values for the $t$-test for $i = 1,\ldots,4$. The linear main effects for $T_{in}$ and $T_{wall}$ are relatively large, -2.77 and 5.450, respectively and their p-values are quite small, 1.59e-08 and 1.543e-22, respectively; therefore $T_{in}$ and $T_{wall}$ are the two most significant factors. $\hat{\beta}_{a2}$ and $\hat{\beta}_{a4}$ have different signs, this implies that $T_{in}$ and $T_{wall}$ have opposite effects on the response. This agrees with the known physics of the problem, i.e., a decrease in $T_{in}$ or an increase in $T_{wall}$ causes an increase in the total rate of steady state heat transfer. As shown in Table 3, the p-values for $\hat{\beta}_{a1}$ and $\hat{\beta}_{a3}$ are quite large. Therefore, $\dot{m}$ and $k$ do not have significant linear main effects on the response in this region of the design space.

**Table 3. Results for $\hat{\boldsymbol{\beta}}_a$**

| | $\hat{\beta}_{a0}$ | $\hat{\beta}_{a1}$ | $\hat{\beta}_{a2}$ | $\hat{\beta}_{a3}$ | $\hat{\beta}_{a4}$ |
|---|---|---|---|---|---|
| Values | 20.606 | 0.409 | -2.77 | 0.673 | 5.450 |
| P-values | | 0.449 | 1.59e-08 | 0.106 | 1.543e-22 |

The maximum likelihood estimators for the correlation parameters $\hat{\boldsymbol{\theta}}_a$ are (1.1780, 0.904, 0.300, 0.01). These values are quite different from each other; therefore different factors affect the correlation of two close point in different scales. Among them, the correlation parameters for $\dot{m}$ and $T_{in}$ are relatively high. The responses of two points, even if there is a small distance between them in the $\dot{m}$-dimension or the $T_{in}$-dimension, may still have a low correlation. Note that $\dot{m}$ does not have a significant linear main effect but

has a large value for its correlation parameter. This implies that the relationship between $\dot{m}$ and the response is nonlinear. This observation may aid our understanding of its physical relationship.

The basic surrogate model is consistent with our background knowledge of the physics of the problem. In general, one would expect the mass flowrate, $\dot{m}$, the temperature of the heat source, $T_{wall}$, and the thermal conductivity of the material, $k$, to have positive linear main effects on the total rate of steady state heat transfer; on the other hand, $T_{in}$ should have a negative linear main effect. The signs of the linear main effects in Table 3 correspond to our expectations. Also, one would expect the temperatures, $T_{in}$ and $T_{wall}$, to have more significant linear main effects on the response than the mass flowrate, $\dot{m}$, or the thermal conductivity, $k$—two factors that have much more complex relationships with the response via the Reynold's number and the temperature gradients throughout the structure, respectively. Their linear main effects are dominated in this region of the design space by the strong linear relationship between the temperatures and the response. However, we might expect them to have significant nonlinear relationships with the response, and we observe this for the mass flowrate, $\dot{m}$.

### 3.4 Using Detailed Simulation Data to Adjust the Base Surrogate Model

Both $y_d$ and $y_a$ are generated for 22 factor level combinations. Fig. 4 presents a plot of $y_d$ vs. $y_a$ for these 44 experiments. It is clear that the detailed simulation and the approximate simulation values are quite different. Some detailed simulation values are higher than approximate simulation values, while some are lower. This demonstrates the need for modeling $\rho(\mathbf{x})$ as a function of $\mathbf{x}$ in Eq. (13).

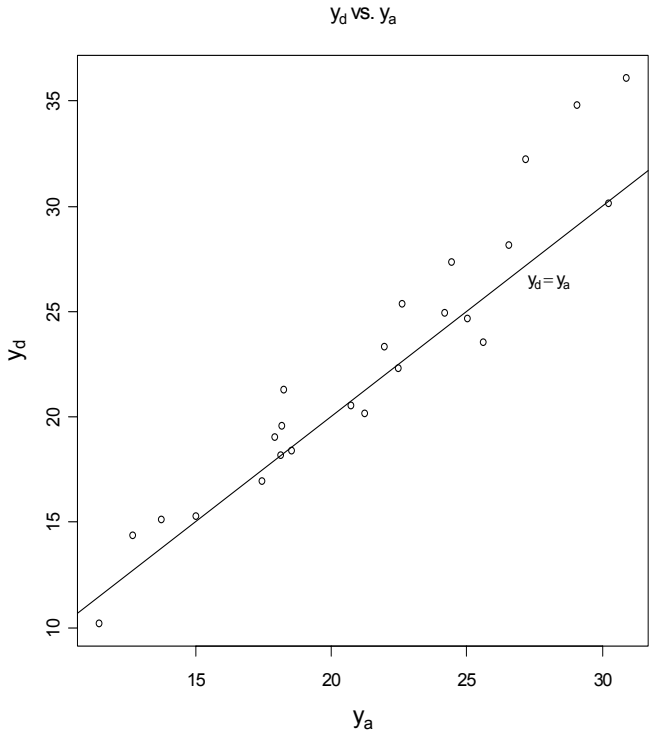


**Fig. 4** $y_d$ **vs.** $y_a$ **for the same design values, where the straight line is** $y_d = y_a$

Next we use the more reliable detailed simulation output, $y_d(\mathbf{x}_i)$, to adjust the fitted model of $y_a(\mathbf{x}_i)$, as described in Section 2.4. Overall, we have a good fit for the adjusted model as $\hat{\sigma}_\delta^2$ has a small value of 0.00515. For the scale adjustment term $\rho(\mathbf{x})$ the parameter estimates are $(\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4) =$ (1.130, 0.090, -0.032, 0.004, -0.012). Among these estimates, the coefficients for $\dot{m}$ and $T_{in}$ are relatively large with significant p-values of 2.165e-23 and 3.839e-13. For the location adjustment term $\delta(\mathbf{x})$, the results are $\hat{\delta}_0 = -0.690$ with the p-value 0.0102 and $\hat{\boldsymbol{\theta}}_\delta =$ (0.173, 0.176, 0.01, 3.66). In Fig. 5, plots of $\hat{\delta}$ vs. different pairs of variables are plotted. In each plot, a 40*40 equally-spaced grid is chosen for the two variables used for plotting and the values of the other two remaining variables are fixed at their mean values**.**

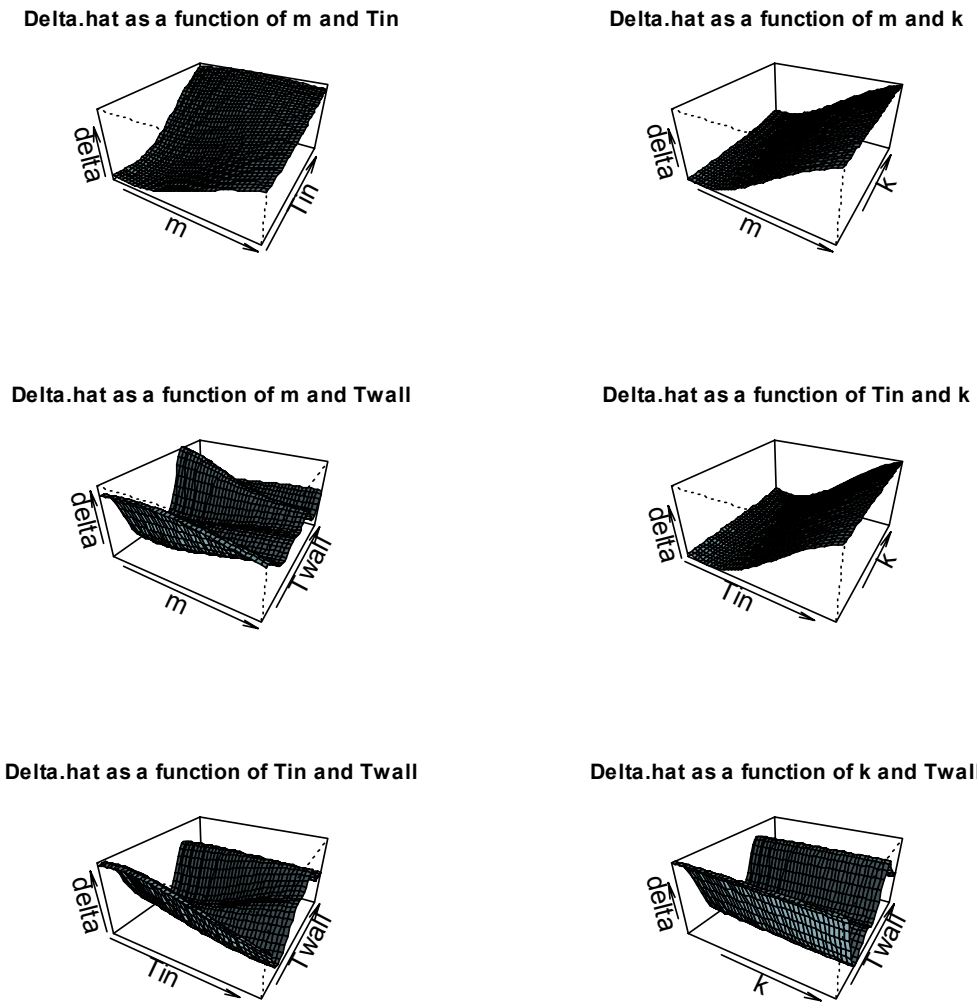


**Fig. 5 $\hat{\delta}$ for different pairs of factors**

Finally, for a new input $\mathbf{x}^*$ we can create the final surrogate model:

$$\hat{y}_d(\mathbf{x}^*) = \hat{\rho}(\mathbf{x}^*)\hat{y}_a(\mathbf{x}^*) + \hat{\delta}(\mathbf{x}^*), \tag{19}$$

where $\hat{\rho}(\mathbf{x}^*) = 1.130 + 0.090x^*_1 - 0.032x^*_2 + 0.004x^*_3 - 0.012x^*_4$. $\hat{y}_a(\cdot)$ is the BULP of $y_a(\cdot)$ as described in Eq. (10) and $\hat{\delta}(\cdot)$ is the BLUP of $\delta(\cdot)$ in Eq. (17).

### 3.5 Validation of the Final Surrogate Model

In order to test and validate the method, 14 additional experiments are performed. In order to test the prediction performance of the final surrogate model in a larger space, some of these validation experiments are selected beyond the original ranges shown in Table 1. For each experimental point, both detailed and approximate simulations are performed. Table 4 lists the factor levels for these experiments, the $y_a$ and $y_d$ values, the predicted $\hat{y}_d$ obtained using Eq. (18) and the predicted $\hat{y}_a$ obtained using Eq. (10).

**Table 4. Additional simulations for validation**

| Run | $\dot{m}$ (kg/s) | $T_{in}$ (K) | $k$ (W/mK) | $T_{wall}$ (K) | $y_d$ | $\hat{y}_d$ | $\hat{y}_a$ | $y_a$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.00050 | 293.15 | 362.73 | 393.15 | 25.82 | 23.85 | 26.96 | 27.24 |
| 2 | 0.00055 | 315 | 310 | 365 | 7.48 | 10.31 | 12.44 | 7.02 |
| 3 | 0.00056 | 277.01 | 354.98 | 374 | 19.77 | 26.02 | 26.38 | 25.53 |
| 4 | 0.00062 | 275 | 225 | 340 | 18.78 | 16.64 | 16.14 | 16.40 |
| 5 | 0.00068 | 313.28 | 259.12 | 350 | 4.55 | 6.44 | 7.32 | 10.23 |
| 6 | 0.00070 | 288.15 | 300 | 400 | 34.45 | 31.93 | 30.97 | 30.90 |
| 7 | 0.00078 | 292.73 | 267.84 | 369 | 21.97 | 23.70 | 22.01 | 20.92 |
| 8 | 0.00080 | 303.15 | 250 | 350 | 14.83 | 6.34 | 6.45 | 13.08 |
| 9 | 0.00085 | 270 | 325 | 385 | 32.85 | 37.88 | 31.34 | 31.14 |
| 10 | 0.00085 | 301.31 | 317.85 | 341 | 11.92 | 12.99 | 11.94 | 11.30 |
| 11 | 0.00091 | 248.87 | 206.74 | 398 | 47.05 | 51.77 | 39.63 | 36.56 |
| 12 | 0.00094 | 271.32 | 362.73 | 400 | 42.93 | 44.97 | 35.63 | 35.53 |
| 13 | 0.00095 | 280 | 270 | 330 | 17.41 | 16.82 | 13.51 | 13.54 |
| 14 | 0.00100 | 293.15 | 202.4 | 373.15 | 22.89 | 25.74 | 21.1 | 21.60 |

Root-mean-square-errors (RMSE) are computed to assess prediction performance. Here we present three different comparisons. The first is a comparison between predictions with the final surrogate model in Eq. (16) and detailed simulation data. The second is a comparison between predictions using the base surrogate model in Eq. (10) and the detailed simulation data, and the third is a comparison between approximate and detailed simulation data.

$$RMSE_1 = \sqrt{\frac{\sum_{j=1}^{14} (\hat{y}_d(\mathbf{x}_j) - y_d(\mathbf{x}_j))^2}{14}} = 3.795,$$

$$RMSE_2 = \sqrt{\frac{\sum_{j=1}^{14} (\hat{y}_a(\mathbf{x}_j) - y_d(\mathbf{x}_j))^2}{14}} = 4.595,$$

and

$$RMSE_3 = \sqrt{\frac{\sum_{j=1}^{14} (y_a(\mathbf{x}_j) - y_d(\mathbf{x}_j))^2}{14}} = 4.431.$$

The proposed method provides a significant improvement in terms of prediction accuracy. The RMSE between $\hat{y}_d$ and $y_d$ is 3.795, which is 14% smaller than the RMSE (4.430) between $y_a$ and $y_d$, and 17% smaller than the RMSE (4.595) between $\hat{y}_a$ and $y_d$ given in Table 2. The difference between these RMSE's is statistically significant. Fig. 5 shows the nonlinear nature of the location adjustment in our procedure. The flexible scale-location adjustment is capable of refining the base surrogate model and obtaining a more accurate surrogate model.

At this point, it is important to determine whether the improvement in prediction accuracy realized with the proposed method justifies the computational expense of building the final surrogate model. Whereas the RMSE of the base surrogate model, $\hat{y}_a$, is 17% larger than the RMSE of the final surrogate model, $\hat{y}_d$, the cost of building the base surrogate model is essentially negligible compared with the cost of building the final surrogate model, requiring minutes versus days of computing time to obtain the approximate and detailed experimental data reported in Table 2. Based on this comparison, a designer may conclude that the improvement in prediction accuracy is not sufficient to justify the increased computational expense of the proposed method. However, the comparison is misleading. In typical

engineering applications, a designer would not rely exclusively on data from an uncalibrated approximate model. Because the accuracy of an approximate model is not known *a priori* in an engineering application, data from detailed simulations or physical experiments are typically conducted throughout the region of interest for validation and calibration. If a number of detailed experiments are conducted anyway, the proposed method is both effective and efficient. By gathering only a few additional detailed simulation data points (beyond the number typically required for validating the approximate model) and by strategically choosing their locations, it is possible to assess the accuracy of an approximate model *and* reduce its predication error using the proposed method.

### 3.6 Maximize the Total Rate of Steady State Heat Transfer

Note that one of the design objectives is to maximize the total heat transfer rate. The ranges of design variables are listed in Table 1. Table 5 contains the maximization results of $\hat{y}_d(x)$ over the ranges. All the optimal values of four design variables are attained at the boundaries of the ranges. These results are not surprising. For this problem we know that as $\dot{m}$ increases, $T_{in}$ decreases, $k$ increases, or $T_{wall}$ increases, the heat transfer rate increases. The maximum value of $\hat{y}_d(\mathbf{x})$, 46.93 is larger than the $y_d$ values given in Tables 2 and 4, except for run 11 in Table 4. This can be explained by noting that the design values for this run are outside the ranges defined in Table 1.

**Table 5. Maximizing $\hat{y}_d(x)$ over the acceptable ranges**

| $\dot{m}$ (kg/s) | $T_{in}$ (K) | $k$ (W/mK) | $T_{wall}$ (K) | $\hat{y}_d(\mathbf{x})$ |
|---|---|---|---|---|
| 0.001 | 270.00 | 360.0 | 400 | 46.93 |

## 4 Closure

In summary, we have presented an approach for building surrogate models based on data from both detailed and approximate simulations. From a design perspective, surrogate models reduce the computational cost of exploring large regions of the design space by replacing detailed simulations.

However, there is still substantial computational cost involved in using detailed simulations to generate data from which surrogate models can be built. Using the approach presented in this paper, it is possible to build accurate surrogate models by supplementing relatively few data points from computationally expensive detailed simulations with abundant data from inexpensive but less accurate approximate simulations. Thus, it is possible to explore a design space with surrogate models that are nearly as accurate but much less computationally expensive than similar surrogate models based exclusively on detailed simulations.

As illustrated with the heat exchanger design example, surrogate models based on approximate and detailed simulations are more accurate than surrogate models based on approximate simulations alone for predicting responses in design regions of interest. Furthermore, the computational effort required to construct a final surrogate model is limited compared with the cost of using detailed simulations exclusively to support the design process. In addition, surrogate models can be adaptively modified when new simulation results are available. The updating of surrogate models only involves refitting the model with old and new data, which requires negligible computational cost.

Another computational advantage of the proposed method is that it does not suffer from the "curse of dimensionality;" therefore, it can be scaled easily to problems with large numbers of design variables. Since only linear terms and constant terms are assumed in modeling the mean part of the process and the dependence among different design points is represented by a correlation function, the number of required design points increases linearly with the number of variables. As a result, for the proposed method the designer does not need to restrict the number of design variables as severely.

In the proposed approach, the final surrogate model interpolates the observed detailed simulation data. This feature is desirable for modeling data from computer experiments, but it implies that the fit of the model is relatively sensitive to individual observations. In some applications, the data may be noisy and include outliers, in contrast to the deterministic, noise-free results encountered for the example problem. If outliers exist in the data, it is important to detect them and remove them from the data set. A simple cross validation procedure can be used to check for outliers. The basic idea is to remove one observation, say ($x_i$,

$y_i$), and use the remaining data $y_{-i}$ to predict it. The quantity $l_i = \frac{y_i - y_{-i}}{s(x_{-i})}$ may be used to check for outliers, where $s(x_{-i})$ is the residual for $y_{-i}$ [48]. The larger $|l_i|$ is, the more likely ($x_i$, $y_i$) is an outlier.

This approach is broadly applicable to examples and phenomena from structural, electrical, financial, and other domains. The primary assumptions are that multiple models or data sources are available and that one model or data source is generally more accurate than the other(s). The method is presented currently to integrate simulation models at only two levels, namely, detailed and approximate. Work is in progress to extend the method for more than two levels of models or data sources. The models may correspond to different physics-based models or approximations of a problem (e.g., Euler Equations vs. Navier-Stokes, etc.).


## Acknowledgments

C.C. Seepersad is sponsored by a Doctoral Fellowship from the Fannie and John Hertz Foundation. Financial support from NSF DMI-0085136 and AFOSR MURI 1606U81 is gratefully acknowledged. Qian and Wu are supported by NSF DMS 0072489. We are grateful to D. L. McDowell and B. M. Dempsey for their assistance and insight about the cellular material example and for the finite difference software used to generate the approximate simulation data. The authors appreciate the efforts of Thang Nguyen and Olu Ogunsanya who ran many of the computer experiments.


## References


1. Wu, C.F.J. and Hamada, M., 2000, *Experiments, Planning, Analysis and Parameter Design Optimization*. New York: John Wiley and Sons.
2. Montgomery, D.C., 1997, *Design and Analysis of Experiments, Fourth Edition*. New York: John Wiley & Sons.
3. Chen, W., Allen, J.K., Tsui, K.-L. and Mistree, F., 1996, "A Procedure for Robust Design". *ASME Journal of Mechanical Design*, Vol. 118(4): p. 478-485.

4. Seepersad, C.C., Dempsey, B.M., Allen, J.K., Mistree, F. and McDowell, D.L., 2004, "Design of Multifunctional Honeycomb Materials". *AIAA Journal*, Vol. 42(5): p. 1025-1033.
5. Michelena, N., Park, H. and Papalambros, P. 2002, "Convergence Properties of Analytical Target Cascading" . *AIAA Journal*, Vol. 41(5), p. 897-905.
6. Myers, R.H. and Montgomery, D.C., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. New York, NY: John Wiley and Sons.
7. Matherton, G., 1963, "Principles of Geostatistics". *Economic Geology*, Vol. 58: p. 1246-1266.
8. Cressie, N., 1988, "Geostatistics". *American Statistician*, Vol. 43(4): p. 197-202.
9. Laslett, G.M., 1994, "Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications". *Journal of the American Statistical Association*, Vol. 89: p. 391-400.
10. Simpson, T.W., Peplinski, J.D., Koch, P.N. and Allen, J.K., 2001, "Metamodels for Computer-Based Engineering Design". *Engineering with Computers: An International Journal for Simulation-Based Engineering (Special issue in honor of Professor S.J. Fevnes)*, Vol. 17: p. 129-150.
11. Wujek, B.A. and Renaud, J.E., 1998, "New Adaptive Move-Limit Management Strategy for Approximate Optimization. Part 1". *AIAA Journal*, Vol. 36(10): p. 1911-1921.
12. Wujek, B.A. and Renaud, J.E., 1998, "New Adaptive Move-Limit Management Strategy for Approximate Optimization: Part 2". *AIAA Journal*, Vol. 36(10): p. 1922-1934.
13. Rodriguez, J.F., Perez, V.M., Padmanabhan, D. and Renaud, J.E., 2001, "Sequential Approximate Optimization Using Variable Fidelity Response Surface Approximations". *Structural and Multidisciplinary Optimization*, Vol. 22: p. 24-34.
14. Akexandrov, N., Dennis, J.E.J., Lewis, R.M. and Torczon, V., 1998, "A Trust Region Framework for Managing the Use of Approximation Models in Optimization". *Structural Optimization*, Vol. 15(1): p. 16-23.
15. Chen, W., Allen, J.K., Schrage, D.P. and Mistree, F., 1997, "Statistical Experimentation Methods for Achieving Affordable Concurrent Systems Design". *AIAA Journal*, Vol. 35(5): p. 893-900.
16. Toropov, V., van Keulen, F., Markine, V. and de Doer, H. 1996, "Refinements in the Multi-Point Approximation Method to Reduce the Effects of Noisy Structural Response". in *6th*

*AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. Bellevue, WA: AIAA, Vol. 2, p. 941-951.

17. Wang, G., Dong, Z. and Atchison, P., 2001, "Adaptive Response Surface Method - a Global Optimization Scheme for Computation-Intensive Design Problems". *Engineering Optimization*, Vol. 33(6): p. 707-734.
18. Wang, G., 2003, "Adaptive Response Surface Method Using Inherited Latin Hypercube Designs". *AMSE Journal of Mechanical Design*, Vol. 125(2): p. 210-220.
19. Farhang-Mehr, A., Azarm, S., 2003, "An Information-Theoretic Entropy Metric for Assessing Multiobjective Optimization Solution Set Quality," *Journal of Mechanical Design*, Vol. 125(4), p. 655-663
20. Lin, Y., Mistree, F. and Allen, J.K. 2004, "A Sequential Exploratory Experimental Design Method: Development of Appropriate Empirical Models in Design". in *ASME Design Engineering Technical Conferences - Design Automation*. Salt Lake City, UT: ASME, Paper Number DETC2004/DAC-57527.
21. Wang, G.G. and Simpson, T.W., 2004, "Fuzzy Clustering Based Hierarchical Metamodeling for Design Space Reduction and Optimization". *Engineering Optimization*, Vol. 36(3): p. 313-335.
22. Box, G.E.P. and Draper, N.R., 1969, *Evolutionary Operation: A Statistical Method for Process Management*. New York, NY: John Wiley and Sons.
23. Koch, P.N., Simspon, T.W., Allen, J.K. and Mistree, F., 1999, "Statistical Approximations for Multidisciplinary Optimization: The Problem of Size". *Journal of Aircraft: Special Multidisciplinary Design Optimiztion Issue*, Vol. 36(1): p. 275-286.
24. Watson, G.S., 1961, "A Study of the Group Screening Method". *Technometrics*, 3(3): p. 371-388.
25. Bettonvil, B., 1990, *Detection of Important Factors by Sequential Bifurcation*. Tilburg, The Netherlands: Tilburg University Press.
26. Bettonvil, B. and Kleijnen, J.J.C., 1996, "Searching for Important Factors in Simulation Models with Many Factors: Sequential Bifurcation". *European Journal of Operational Research*, 1996. p. 180-194.
27. Wu, C.J.F., 1993, "Construction of Supersaturated Designs through Partially Aliased Interactions",. *Biometrika*, Vol. 80(3): p. 661-669.

28. Holcomb, D.R., Montgomery, D.C. and Carlyle, W.M., 2003, "Analysis of Supersaturated Designs". *Journal of Quality Technology*, Vol. 35(1): p. 13-27.
29. Qian, A., Seepersad, C.C., Joseph, V.R., Wu, C.F.J. and Allen, J.K. 2004, "Building Surrogate Models Based on Detailed and Approximate Simulations". in *ASME Design Engineering Technical Conferences - Design Automation*. Salt Lake City, UT: ASME, Paper Number DEC2004/ 57486.
30. Osio, I.C. and Amon, C.H., 1996, "An Engineering Design Methodology with Multistage Bayesian Surroagtes and Optimal Sampling". *Research in Engineering Design*, Vol. 8: p. 189-206.
31. Pacheco, J.E., Amon, C.H. and Finger, S., 2003, "Bayesian Surrogates Applied to Conceptual Stages of the Engineering Design Process". *Journal of Mechanical Design*, Vol. 125: p. 664-672.
32. Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P., 1989, "Design and Analysis of Computer Experiments". *Statistical Science*, Vol. 4: p. 409-435.
33. Santner, T.J., Williams, B.J. and Notz, W.I., 2003, *The Design and Analysis of Computer Experiments*. New York: Springer.
34. Simpson, T.W., Mauery, T.M., Korte, J.J. and Mistree, F., 2001, "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization". *American Institute of Aeronautics and Astronautics Journal*, Vol. 39(12): p. 2233-2241.
35. Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J. and Morris, M.D., 1992, "Screening, Predicting and Computer Experiments". *Technometrics*, Vol. 34: p. 15-25.
36. Computing, R.P.f.S., A Statistical Computer Environment Http://Www.R-Project.Org.
37. Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C., 1995, "A Limited Memory Algorithm for Bound Constrained Optimization". *SIAM Journal Scientific Computing*, Vol.. 16, p. 1190-1208.
38. Handcock, M.S. and Stein, M.L., 1993, "A Bayesian Analysis of Kriging". *Technometrics*, Vol. 35(4), p. 403-410.
39. Handcock, M.S. and Wallis, J.R., 1994, "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields. *Jounral of the American Statistical Association*", Vol. 89, p. 386-378.
40. Currin, C., Mitchell, T.J., Morris, M. and Ylvisaker, D., 1991, "Bayesian Predications of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments". *Journal of the American Statistical Association*, Vol. 86, p. 953-963.

41. Cochran, J.K., Lee, K.J., McDowell, D.L. and Sanders, T.H. 2000, "Low Density Monolithic Honeycombs by Thermal Chemical Processing." in *Proceedings of the 4th Conference on Aerospace Materials, Processes, and Environmental Technology*. Huntsville, AL.
42. Gibson, L.J. and Ashby, M.F., 1997, *Cellular Solids: Structure and Properties*. Cambridge, UK: Cambridge University Press.
43. Hayes, A.M., Wang, A., Dempsey, B.M. and McDowell, D.L. 2001, "Mechanics of Linear Cellular Alloys". in *Proceedings of IMECE, International Mechanical Engineering Congress and Exposition*. New York, NY.
44. Evans, A.G., Hutchinson, J.W., Fleck, N.A., Ashby, M.F. and Wadley, H.N.G., 2001, "The Topological Design of Multifunctional Cellular Materials". *Progress in Materials Science*, 46(3-4): p. 309-327.
45. Fluent, I., *Fluent*. Release 5.5.14 (3d, segregated, laminar), 1998.
46. Incropera, F.P. and DeWitt, D.P., 1996, *Fundamentals of Heat and Mass Transfer*. 3rd Edition ed. New York: John Wiley & Sons.
47. Dempsey, B.M., 2002, *Thermal Properties of Linear Cellular Alloys*, MS Thesis, G..W. Woodruff School of Mechanical Engineering. Georgia Institute of Technology: Atlanta, GA.
48. Cook, D.R. and Weisberg, S., 1994, *An Introduction to Regression Graphics*. New York: John Wiley and Sons.