ABSTRACT
Computer models and simulations are essential system design tools that allow for improved decision making and cost reductions during all phases of the design process. However, the most accurate models tend to be computationally expensive and can therefore only be used sporadically. Consequently, designers are often forced to choose between exploring many design alternatives with less accurate, inexpensive models and evaluating fewer alternatives with the most accurate models. To achieve both broad exploration of the design space and accurate determination of the best alternatives, surrogate modeling and variable accuracy modeling are gaining in popularity. A surrogate model is a mathematically tractable approximation of a more expensive model based on a limited sampling of that model. Variable accuracy modeling involves a collection of different models of the same system with different accuracies and computational costs. We hypothesize that designers can determine the best solutions more efficiently using surrogate and variable accuracy models. This hypothesis is based on the observation that very poor solutions can be eliminated inexpensively by using only less accurate models. The most accurate models are then reserved for discerning the best solution from the set of good solutions.

In this paper, a new approach for global optimization is introduced, which uses variable accuracy models in conjunction with a kriging surrogate model and a sequential sampling strategy based on a Value of Information (VOI) metric. There are two main contributions. The first is a novel surrogate modeling method that accommodates data from any number of different models of varying accuracy and cost. The proposed surrogate model is Gaussian process-based, much like classic kriging modeling approaches. However, in this new approach, the error between the model output and the unknown truth (the real world process) is explicitly accounted for. When variable accuracy data is used, the resulting response surface does not interpolate the data points but provides an approximate fit giving the most weight to the most accurate data. The second contribution is a new method for sequential sampling. Information from the current surrogate model is combined with the underlying variable accuracy models’ cost and accuracy to determine where best to sample next using the VOI metric. This metric is used to mathematically determine where next to sample and with which model. In this manner, the cost of further analysis is explicitly taken into account during the optimization process.

INTRODUCTION
Systems design problems are often complex, involving many interactions between multiple subsystems [16, 36]. The design space for such problems is large, from initial exploration of plausible architectures through refinement of sizing parameters and detailed design. As a result, modeling, simulation, and optimization have become increasingly vital to engineering design and decision making. While no model is perfectly accurate, models are often the only option for design space exploration and informed decision making because they
are significantly cheaper than physical experiments and prototypes. While model accuracy is improving due to faster processors and more sophisticated modeling software, added accuracy has come at a cost. Therefore, it is not generally pragmatic or even possible to explore an entire design space at a high level of detail due to time and computational costs incurred. To counter this problem, methods have been proposed for reducing the number of design variables by using screening methods [39] or by performing a sensitivity analysis [7]. Alternatively, methods have been developed for approximating an otherwise computationally intractable functional relationship by using a simplified or surrogate model [7, 9, 19-20, 29, 34-35, 38-39, 42]. By reducing the number of design variables or approximating expensive functions with cheaper surrogates, optimizations and design space explorations can be performed at a reasonable cost.

Design problems are frequently framed as optimization problems; the designer is seeking the best possible artifact to meet her needs. For engineering applications, however, simply finding the mathematical optimum of a design optimization problem is not the primary objective. Even if sufficient computing resources are available, it is generally impossible to mathematically prove global optimality for complex, black-box simulation models. Therefore, what is actually desired is not optimality, but a sufficiently good solution that can be achieved at a reasonable design process cost. For example, an electronic widget that is ‘optimally’ designed may have no real value if it takes two years to find that solution and implement it; by then a competitor may have already produced another widget or consumer demands may have shifted. In such a case, the cost of the design process must be considered because it may greatly affect the utility of the final widget.

To that end, this paper is written from a Rational Design perspective, an extension of Decision-Based Design (DBD) [17, 25, 41]. In DBD, it is assumed that decisions are best made using mathematically sound methods derived from decision theory; in particular, design decisions are made axiomatically using a utility function to compare across alternatives. There are well-established methodologies for eliciting a utility function to quantify and combine attributes that are material to the decision maker [6, 30]. While such utility functions may provide a mathematical means of comparing the relative profitability or goodness of a particular design artifact, the costs incurred during the design process are generally neglected. In Rational Design Theory (RDT), however, the costs of the process and analyses are included in the utility calculations and modeled explicitly [40]. This approach provides a mathematically sound mechanism for considering the overall utility of the artifacts and the cost of the design process. Specifically, RDT provides mathematical justification for using the available models and analyses in a cost effective way.

BACKGROUND AND PRIOR WORK

KRIGING MODELING APPROACHES

Kriging modeling [7-8, 11, 15, 26, 34-35, 39, 42] is an interpolation technique that has its origins in geostatistics literature [8] and is now very prolific in engineering domains. Kriging modeling was first applied to computer experiments by Sacks et al [35]. In this approach, kriging models are fit to data from a deterministic computer model under the assumption that interpolation is a desirable characteristic in the absence of measurement noise. In many cases, interpolating the sample sites is an appealing attribute, but this means that kriging predictors do not explicitly allow for model error. That is, it is assumed that the underlying model is perfectly accurate; the error at the observation sites where the original model has been sampled is identically zero. Under this assumption, it is logical to interpolate the given data.

Kriging models have additional appealing properties that cause them to be very prolific in the engineering literature. Mathematically, kriging models are rooted in statistics and consist of a sum of a regression model and a realization of a zero-mean Gaussian process model. This model construction allows for a lot of flexibility in model properties and fits, but like many practical engineering design spaces, kriging models tend to be continuous and smooth. In addition, because kriging models are statistical in nature, the hyperparameters for the predictor can be determined without user input by using maximum likelihood estimation. Finally, kriging models also have the advantage of providing an equation for mean-squared error (MSE) of the predictor. While the model being approximated is assumed to have no error, error is nonzero between the kriging predictor and the underlying model being sampled, particularly at points that are far from any sample sites. This predictor error is quantified as MSE, and having an explicit equation for it allows the user to gain insight about the model’s reliability anywhere in the design space.

Another limitation imposed by kriging is that interpolating the sample points is only desirable if the design sites are only drawn from one coherent set of data from one model. Combining data from multiple models of varying accuracy and interpolating the design sites would not result in a meaningful fit. More importantly, having multiple data points at the same input, even if they came from different models, would cause a classical kriging fit to fail due to a mathematical singularity. To handle data from multiple models with uncertainty, a modified approach will be required.

One modified kriging algorithm has been developed to accommodate $n$ levels of fidelity data in the model [18, 22]; however, the design sites for the higher fidelity models must be a subset of the design sites at which the lowest fidelity model is sampled. Additionally, this version of kriging modeling for multi-accuracy data assumes that the different models are correlated, which requires estimation of a greater number of hyperparameters than the approach presented in this paper.

Another similar but not identical kriging formulation is stochastic kriging, as presented in [4]. While this model allows
for model error, the error is due to the stochastic nature of discrete simulations. These simulations are not deterministic, so an error term is added at the design sites. This is different from model inadequacy, which is addressed in this paper. Model inadequacy describes a deterministic simulation which does not identically reflect reality, and this error is treated differently in the Gaussian process modeling formulation presented in this paper.

**SAMPLING APPROACHES FOR RESPONSE SURFACE MODELING**

Surrogate models of all varieties are fit to some sample sites which are determined using the original underlying model. Generally, there must be a sufficient number of samples to have an accurate surrogate model, but over-sampling results in excess cost. There are two general strategies for sampling. One option is that a surrogate model can be fit to a predetermined set of samples selected using either a design of experiments or other selection criteria for a given number of samples. Alternatively, design sites can be added incrementally with the model being dynamically updated as new samples are added. If the intent is to get a reasonably accurate view of the entire design space, then a global, fixed sampling approach is likely to be a good choice. On the other hand, if optimization is the primary concern, it is not economical to sample the entire design space evenly or even to achieve uniform accuracy throughout the space; for optimization, high accuracy is only needed near the optimum. In this context, an incremental sampling approach is more likely to be appropriate. Both categories of approaches have their merits and have been used in conjunction with kriging models.

**Fixed Sampling Approaches**

Fixed sampling approaches are generally used when it is necessary to model the entire design space with a relatively consistent degree of accuracy; the objective is to visualize a complex function while saving cost by using a surrogate. In this case, one common approach is to use a space-filling design of experiments to select the design sites. Latin hypercube sampling (LHS) [27] is a stratified sampling technique that is widely used throughout the literature in conjunction with kriging models [42]. While LHS ensures that the model is sampled throughout the range of all the input variables, it has been shown that LHS is often too sparse for capturing all of the details of complex models [35]. While simply increasing the number of samples may reduce the error of the kriging model, adding samples in areas in which the model is already sufficiently accurate is a waste of resources. Because kriging models include a closed form solution for MSE, this parameter can be used to determine where the model is most inaccurate. This leads to some different types of designs for fitting a kriging model where the samples are not equally spaced, but rather, spaced in a way that guarantees optimal accuracy with respect to a particular metric for a fixed number of samples.

In [35], some different metrics are presented for optimizing design site selection assuming a fixed number of design sites. These metrics are appropriate for kriging modeling in that they rely on the estimate of MSE given by the kriging model. The first metric introduced is Integrated Mean Squared Error (IMSE). In this method, the design sites are selected such that the integral of the MSE function throughout the design space is minimized. This metric is appropriate when global accuracy is desired. The second metric presented is Maximum Mean Squared Error (MMSE), where the design sites are selected such that the maximum of the MSE in the design space is minimized. While this metric is still appropriate for global accuracy, the emphasis of MMSE is different from that of IMSE; for MMSE the maximum MSE present at any point in the design space is minimized, whereas IMSE does not bound the error at any one point, but rather seeks to minimize the total amount of error in the design space.

While all of these methods are appropriate in a global exploration context, none of them is ideal for design optimization from an RDT perspective. First, the number of design sites must be pre-specified without any mathematical justification of the cost incurred and the quality of the outcome. Additionally, none of these methods allow for variable accuracy throughout the space; no resources are channeled toward finding the optimum. From an RDT perspective, it only makes sense to allocate significant resources in promising regions. To accomplish this, it is important to examine incremental approaches for adding design points so that the current state of knowledge dictates where it is most valuable to expend additional computational effort.

**Incremental Sampling Approaches**

Metrics for selecting design sites and adding them to the design space incrementally have been presented in the literature, primarily in situations where the emphasis is on optimization rather than global exploration. However, the metric used to select design points is independent of the decision to add samples incrementally. For example, the same criteria presented in the previous section (IMSE, MMSE, posterior entropy) can be employed in incremental schemes. The difference is that rather than deciding how many points to sample a priori and running an optimization to select those points, the same metrics can be used to optimize the selection of only one point or a few points at a time. Then, the surface is updated before adding additional points and before selecting which new point(s) to add in the next iteration.

In the global optimization literature, several metrics are presented for selecting the next design sites based on the current fit [19]. While these metrics are all used in incremental sampling schemes, the emphasis for many of these metrics is strictly optimization. Therefore, rather than focusing on the overall accuracy of the entire space, these metrics are designed to favor finding the best solution. One such metric is maximizing the Probability of Improvement. The PI metric is a function for calculating the probability of improving over the current best design site by some pre-specified target value. The target value is set as a percentage improvement over the current best, and it is also possible to use multiple target values. Conceptually, this is an appropriate metric for global optimization; samples are added if and only if they have the
highest probability of resulting in improvement over the current best point based on the current kriging model. Additionally, sampling stops when the probability of improving by some percentage is equal to zero. However, one disadvantage of this approach is that the target value must be specified by the user. Also, while the stopping criterion is intuitive, the cost of performing additional function evaluations is not taken into account. It is likely that situations would arise when the cost of running an additional analysis would outweigh the improvement that could be achieved over the current best, or improvement could stop while it is still economical to evaluate additional points.

Another incremental metric presented in the global optimization literature is Expected Improvement [12, 18-20]. This metric is similar to the PI approach, but in this metric the expected amount of improvement over the current best is maximized. Rather than targeting a design site that is most likely to yield any improvement, the design site with the greatest amount of expected improvement is selected. Again, this criterion is appropriate in global optimization, and the focus is not on global exploration, but on finding the global optimum. There is an intuitive stopping criterion as well; stop when the expected improvement is smaller than some constant. There is also a multi-point version of this criterion [12] in which multiple design sites are selected for evaluation in only one iteration, which reduces the cost of fitting a kriging surface after each new point is added. The shortcomings of this metric are that uncertainty and model accuracy are not taken into account, nor is the cost of analyzing the new design site.

There is also a modified expected improvement metric that does account for the cost of additional function evaluations, and this metric is presented in conjunction with a multi-fidelity sequential kriging optimization in [18]. This modified expected improvement metric accounts for cost and includes a number of user defined parameters that can affect the searching behaviors and the stopping criteria. While much of the same data is incorporated in both the proposed VOI metric and the modified expected improvement metric (risk preferences, uncertainty, etc.) this approach is not equivalent to the VOI formulation presented in this paper. The costs are weights with respect to improvement that must be defined by the user, and this metric does not necessarily stop when the cost of additional simulation is equivalent to or outweighs the potential benefit of further search.

VARIABLE ACCURACY MODELING APPROACHES

Variable accuracy modeling is about using multiple models of differing accuracies and costs during optimization in order to allow broad, efficient exploration as well as high accuracy when it is needed. The idea of using variable accuracy models in the optimization process for engineering design has been around for a long time [37]. A simple approach to variable accuracy modeling in optimization is to create feasibility constraints that can be tested quickly using an inexpensive model to eliminate obviously poor solutions before testing the good solutions using a higher accuracy model [13-14, 31]. Other approaches to variable accuracy modeling include the space mapping approach [5] which attempts to create a mapping or correction between a coarse (low accuracy) design space and a fine (high accuracy) design space that will yield the same computational outcome.

Seminal work in variable accuracy modeling has been done by Alexandrov et al. [1-3]. In one example, an optimization is performed using the Euler equation over variable mesh sizes, effectively changing the resolution of the model. In a different example, variable-fidelity physics models are used, where the high-fidelity model is the Navier-Stokes equation and the low fidelity model is the Euler equation. In both examples, the method of correlation is a first order error function in a given trust region using augmented Lagrangian methods, which have been shown to converge to a Karush-Kuhn-Tucker (KKT) feasible point for constrained minimization problems [32]. Using the low fidelity model and this corrective factor, nested optimizations are performed on the low fidelity model, and then the trust region is adjusted based on the performance of the high fidelity model. While this method requires relatively few function calls to the high fidelity model, the method is restricted to derivative based optimization approaches and only two models: one low and one high fidelity. In other similar works [10, 24, 33, 43] a very similar trust region optimization technique is successfully applied using either one of low fidelity model types presented by Alexandrov, or by using a response surface approximation as the low fidelity model.

APPROACH

As discussed previously, models and simulations encountered in engineering disciplines are often computationally expensive to evaluate (e.g., Finite Element Models). Thus, designers are frequently faced with a tradeoff between the number of alternatives that can be evaluated and the accuracy with which they are evaluated. As illustrated in Figure 1, designers are often confronted with a dilemma: Is it better to explore many alternatives with an inexpensive, less accurate model or to explore only a few options with a more expensive, more accurate model?
Using models of varying accuracy [2, 10, 18, 24, 31, 37] enables computational resources to be better channeled toward discerning the best solution from the set of good solutions, rather than performing expensive function evaluations on obviously bad solution alternatives. This hypothesis is based on the observation that high accuracy is only necessary when close to the optimum, as illustrated in Figure 2. There is no need to accurately know how bad a poor design alternative is, so long as we can identify the general direction in which better solutions can be found. Assuming that accurate models require more computing resources than inaccurate models, a very accurate assessment of a bad solution’s inadequacy is a waste of resources. Only when a design alternative is near the optimum is an accurate assessment required; however, a low fidelity model would not allow one to identify the best solution from among these near optimal alternatives. Combining models with multiple accuracies enables global exploration at reasonable cost while still ensuring high accuracy in the neighborhood of the optimum.

Models at multiple levels of accuracy can indeed enable design space exploration in a cost efficient and effective manner, but infrastructure is required for managing information obtained from each of the models. Past approaches to variable accuracy modeling tend to be limited to only two models, and moreover, do not explicitly account for the cost of the analyses used during the process [1, 3, 10, 24, 28, 32-33, 43]. Therefore, what is needed is a method to combine predictions from any number of variable accuracy models so that all of the relevant information can be used and one model does not simply over-ride another. In addition, the available models must be used to make good design decisions in a cost-effective manner; that is, the most accurate and costly models should only be used when it is valuable to do so. Thus, a method for quantifying value and selecting the most valuable analysis action is needed for every step in the exploration process.

To formulate a predictor based on multi-accuracy data, a Gaussian process modeling approach is used to fit a surrogate model to the known points in the design space. The model used is a novel extension of kriging modeling. In kriging modeling, zero uncertainty is assumed at all sample points, resulting in interpolating surrogate models. The proposed approach relaxes this assumption, and allows for model uncertainty. It is assumed that the uncertainty in each model can be adequately characterized by a zero mean Gaussian process with known variance. Instead of interpolating the data, the resulting surrogate model approximates the data weighted by accuracy. That is, the surrogate model will tend closer to points from more accurate models.

For optimization, this surrogate modeling approach is used in conjunction with the Value of Information metric. This metric takes into account the probability of an improved solution over the current best, given the prior mean and variance of the current best sampled data point, the cost of performing an analysis, and the posterior uncertainty after analyzing. This metric provides a mathematically sound mechanism for trading off solution quality, potential for improvement and the cost of performing additional analyses. Additionally, the VOI metric naturally balances global search, based on prior and posterior uncertainty, with local search, based on the likelihood of improving the solution quality and the cost of the analysis. Sometimes a very cheap analysis that reduces the uncertainty in the global space may be of more value than a very costly analysis in the vicinity of the current optimum. Combining the VOI metric for finding the next points for evaluation with the Gaussian process surrogate model for n models, it is possible to navigate a design space of multi-fidelity data in a cost-effective, efficient manner.

**GAUSSIAN PROCESS MODELING FOR MULTIPLE MODELS**

Assume the design space is populated by m design sites in p dimensions obtained by evaluating each of q unique models n_q times at various inputs. The inputs are captured in a matrix $S$ and the outputs are captured in a vector $Y_s$. To simplify the...
math, it is assumed that the output is one-dimensional, restricted to $\mathbb{R}^1$.

$$S = \begin{bmatrix} s_{11}, s_{12}, \ldots, s_{1n}, \ldots, s_{21}, s_{22}, \ldots, s_{2n}, \ldots, s_{q1}, s_{q2}, \ldots, s_{qn} \end{bmatrix}, \quad s_{ij} \in \mathbb{R}^n$$

Following the seminal work of [34], we adopt a model $y$ that consists of a sum of a regression term and a zero mean Gaussian process:

$$Y = \begin{bmatrix} y_{11}, y_{12}, \ldots, y_{1n}, \ldots, y_{21}, y_{22}, \ldots, y_{2n}, \ldots, y_{q1}, y_{q2}, \ldots, y_{qn} \end{bmatrix}, \quad y_{ij} \in \mathbb{R}^1$$

where $f^T(x) \beta$ is the regression function and $z(x)$ is the zero mean Gaussian process. However, unlike most previous kriging modeling approaches, it is assumed that this function is a model of a ‘truth’ that cannot be observed directly. That is, while the outcome of the computer model is deterministic, it is merely an approximation of some true system that is in essence unknown. The difference between the simulation and the true system is characterized as model inadequacy [21]. Therefore, we assume that the observations of the computer model at the design sites can be modeled as follows:

$$y(x) = f^T(x) \beta + z(x), \quad x \in \mathbb{R}^n$$

where $f^T(x) \beta$ and $Z$ are the same as those found in [34], where $f^T \beta$ is a regression model composed of any $p$ chosen functions with $f_j : \mathbb{R}^n \to \mathbb{R}$, and $Z$ is a realization of a zero mean, correlated stationary Gaussian process. This term represents the error between the regression function and the experimental output.

$Z_m$ is the only term in this model formulation that is different from the classic formulation presented in [34]. $Z_m$ is a correlated Gaussian process representing model inadequacy for the different simulation models. That is, for a sample site $s_{ij}$, $Z_m$ represents the model error for model $i$, i.e. the difference between the potentially unknown truth and the $i^{th}$ model. We assume that the model error at site $s_{ij}$ is correlated with the model error at site $s_{ik}$ for the same model $i$, but uncorrelated with the model error for any model $l$, other than $i$. It is also assumed that the model errors, $Z_m$, are uncorrelated with the error, $Z$, in the regression model.

Following the classic kriging derivation, we now consider the linear predictor

$$\hat{y}(x) = c^T Y, \quad c = c(x) \in \mathbb{R}^n$$

Then the error is

$$\hat{y}(x) - y(x) = c^T Y - y(x)$$

$$= c^T (f \beta + Z + Z_m) - (f \beta + z)$$

To keep the predictor unbiased we impose the constraint

$$F^T c - f = 0$$

so that

$$\hat{y}(x) - y(x) = c^T (Z + Z_m) - z$$

The mean squared error (MSE) of the predictor is

$$\varphi(x) = E \left[ (\hat{y}(x) - y(x))^2 \right]$$

$$= E[c^T Z Z^T c + 2c^T ZZ^T c + c^T Z_m Z_m^T c - 2c^T Z z - 2c^T Z_m z + z^2]$$

The goal is to minimize the expected MSE. To do this, the covariance structures must be defined for each of the terms in the expression for MSE.

1) $E[ZZ^T] = \sigma^2 R$

This term is defined the same way as it is in the original derivation. For $Z$, the covariance is assumed to be $\sigma^2 R$ where the correlation $R$ is defined as

$$R_{ij} = R(\theta, s_{ij}, s_{kl})$$

representing the correlation between the Gaussian processes at the design sites. In this notation, $\theta$ is the roughness parameter, characterizing the correlation strength between two points as a function of distance. One of the most commonly used is the Gaussian covariance structure for stationary, one-dimensional correlations. For two vectors $x$ and $x'$ in $p$ dimensions, this correlation structure is defined as

$$R(\theta, x, x') = \prod_{j=1}^{p} \exp(-\theta_j |x_j - x_j'|^2)$$

2) $E[Z Z_m^T] = 0$

Because of the assumption that the model errors, $Z_m$, are uncorrelated with $Z$, the resulting covariance structure is 0.

3) $E[Z_m Z_m^T] = \sum_{i=1}^{q} \sigma_i^2 R_i$

Here, a covariance structure for $Z_m$ must be defined. In this approach, it is assumed that the correlation between the $q$ different models is zero, which is different from previous multi-accuracy kriging modeling approaches [18, 22]. This is based on the observation that the error between one model and the truth has no effect on the error between any other model and the truth. Therefore, all that remains is the covariance structure for samples from the same model. The covariance structure for the individual model $i$ is $\sigma^2_i R_i$, where the correlation structure $R_i$ is defined the same way as $R$ for the Gaussian process $Z$, including only the sample sites from model $i$ and with a separate $n$-dimensional roughness parameter $\theta_i$. While the $\sigma^2_i$ parameter for the covariance for $Z$ is assumed unknown and calculated using Maximum Likelihood Estimation, $\sigma_i^2$ is assumed to be a known parameter characterizing the accuracy of an individual model. A very inaccurate model would be expected to have a larger variance with respect to the truth, while a more accurate model has a smaller variance. For an untried model, the effects of every design site for each of the individual models must be taken into account. Therefore, the final covariance structure for $Z_m$ is
The correlation structure for $Z \cdot z$, where $z$ is the Gaussian process describing the error between the regression model and the values at the design sites, we define $\rho(x)$ as the correlation between an untried point $x$ and the sample sites.

5) $E[Zm] = 0$

$Z_m$ is assumed uncorrelated with $z$, which represents the difference between the regression model and the truth.

6) $E[z^2] = \sigma^2$

This is a property of the Gaussian process $z$.

Thus, the MSE can be rewritten as

$$
\varphi(x) = \sigma^2 c^T R c + c^T \left( \sum_{i=1}^{q} \sigma_i^2 R_i \right) c - 2 \sigma^2 c^T r(x) + \sigma^2
$$

$$
= \sigma^2 \left( 1 + c^T R c - 2c^T r \right) + c^T \left( \sum_{i=1}^{q} \sigma_i^2 R_i \right) c
$$

Recall that

$$
Y_s = \begin{bmatrix}
\text{Model 1} & \text{Model 2} & \cdots & \text{Model } q
\end{bmatrix}
\begin{bmatrix}
y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{q1}, \ldots, y_{qn_q}
\end{bmatrix}
y_{ij} \in \mathbb{R}^1
$$

To solve for the now unknown parameters, begin by minimizing the MSE with respect to $c$.

$$
\min_c \varphi(x) = \sigma^2 \left( 1 + c^T R c - 2c^T r \right) + c^T \left( \sum_{i=1}^{q} \sigma_i^2 R_i \right) c
$$

Subject to the unbiased constraint

$$
F^T c - f = 0
$$

The details of this minimization can be found in the original derivation, but it is important to note that we let

$$
K = \sigma^2 R + \sum_{i=1}^{q} \sigma_i^2 R_i
$$

From the first order necessary conditions for optimality, we get

$$
\begin{bmatrix}
K & F
\end{bmatrix}
\begin{bmatrix}
c
\end{bmatrix}
= \begin{bmatrix}
\sigma^2 R
\end{bmatrix}
$$

Solution:

$$
c = K^{-1} \left( \sigma^2 r + F \frac{\lambda}{2} \right)
$$

$$
\lambda = -2(F^T K^{-1} F)^{-1}(F^T K^{-1} \sigma^2 r - f)
$$

Since $K$ and $K^{-1}$ are symmetric,

$$
\hat{y} = c^T Y = \left( \sigma^2 r + F \frac{\lambda}{2} \right)^T K^{-1} Y
$$

$$
= \sigma^2 r^T K^{-1} Y - (F^T K^{-1} \sigma^2 r - f)^T K^{-1} Y
$$

To further simplify this expression, a generalized Least Squares fit is used. Consider the regression problem $F \beta \cong Y$, where $Y$ represents the realization of a stochastic process.

The generalized least squares fit becomes

$$
(F^T K^{-1} F) \beta^* = F^T K^{-1} Y
$$

$$
\beta^* = (F^T K^{-1} F)^{-1} F^T K^{-1} Y
$$

and the predictor can be rewritten as

$$
\hat{y} = \sigma^2 r^T K^{-1} Y - (F^T K^{-1} \sigma^2 r - f)^T \beta^*
$$

$$
= f^T \beta^* + \sigma^2 r^T K^{-1} (Y - F \beta^*)
$$

$$
\hat{y} = f^T \beta^* + r^T Y
$$

where $Y^* = \sigma^2 K^{-1} (Y - F \beta^*)$.

The MSE then becomes

$$
\varphi(x) = \sigma^2 + u^T (F^T K^{-1} F)^{-1} u - \sigma^2 r^T K^{-1} (\sigma^2 r)
$$

where $u = F K^{-1} (\sigma^2 r) - f$ and $\sigma^2$ is found from maximum likelihood estimation.

To maximize the likelihood:

$$
\max_{\sigma^2, \beta} L = \max_{\sigma^2, \beta, \beta^*} \left( \frac{1}{(2\pi)^{m/2}|K|^{1/2}} e^{-\frac{1}{2} (Y-F \beta^*)^T K^{-1} (Y-F \beta^*)} \right)
$$

Since the individual $\sigma_i$ are assumed to be known as a metric of model accuracy, $L$ is maximized over the remaining unknown parameters. Therefore, we need

$$
\frac{\partial L}{\partial \sigma} = 0, \quad \frac{\partial L}{\partial \beta} = 0, \quad \frac{\partial L}{\partial \beta^*} = 0 \quad \text{for } i = 1, \ldots, q
$$

Because of the complexity of $K$, it is not possible to find a closed form expression for any of the partial derivatives. Therefore, the maximization must be done numerically. To
make the problem more computationally tractable, the natural logarithm of $L$ is maximized,

$$\max_{\sigma, \beta, R_i} \ln(L) = \max_{\sigma, \beta, R_i} \left( -\ln \left( |K|^{1/2} \right) - \frac{1}{2} (Y - F \beta^*)^T K^{-1} (Y - F \beta^*) \right)$$

**VALUE OF INFORMATION AS A SAMPLING CRITERION**

As discussed previously, this paper is written from a Rational Design Theory (RDT) perspective. As such, the assumptions which neglect the cost of the design and decision making process in the more classic Decision-Based Design approach are relaxed. In RDT, the scope is broadened to encompass not just the final artifact and its net utility, but also the resources consumed in making decisions and analyzing options before selecting the final design [40]. In engineering design, a decision maker is often forced to choose an artifact under some amount of uncertainty. Early in the design stages, this uncertainty may be very high; however, in many cases this uncertainty might be reduced by performing additional analyses or simulations on the designs under consideration. To do this efficiently, it is necessary to quantify the amount of uncertainty before and after an analysis is performed and to consider the cost of the. To do this, the Value of Information metric is presented [23].

Assume that there exists some prior belief in the Bayesian sense about a state $x$, $p(x)$. For example, $x$ could represent the state of an artifact given the data that has already been obtained by making some observations. After running some additional analysis about $x$ a message $y$ is obtained. After processing that message, that is, after incorporating that new piece of information into our knowledge state, there is now some posterior belief about $x$ given $y$, or specifically $p(x|y)$. Information theory then gives a mathematical formulation for the ex-post value of a message $y$:

$$v(x, y) = \pi(x, a_y) - \pi(x, a_0)$$

Where $\pi$ is the payoff function, $a_y$ represents the optimal response to the specific message $y$ under the present knowledge $p(x|y)$, and $x$ is the state that occurs after the choice of action is made. The equation above represents the difference between the payoff given choice $a_y$, the choice made given the message $y$, and the payoff that would have resulted under the prior action $a_0$, which would have occurred had the message $y$ not been taken into consideration. This value can be positive or zero; the message $y$ might not lead to a change in action, in which case the outcome would be the same.

This same principle can now be extended to the expected value of a source with information structure $I$, called the incremental expected payoff:

$$V(I) = E_y \ E_{x|y} \pi(x, a_y) - E_x \pi(x, a_0)$$

Here, $V(I)$ represents the maximum payout that could be achieved while incorporating information into the solution. However, we still need to account for the cost of the analysis. This leads to the realized incremental gain:

$$G = v(x, y) - C = \pi(x, a_y) - \pi(x, a_0) - C$$

where $C$ is the cost of obtaining and applying the message $y$, or in this case, the cost of performing an additional analysis. Therefore, if the cost $C$ is equal to or exceeds the value of the source, $V(I)$, then the realized gain is 0 or negative, implying that performing the additional analysis would be a waste of resources. This is precisely the principle applied when considering what point in the design space to analyze next and how much cost can be justified in doing so.

In terms of engineering optimization, this is a novel approach. The value of information gives a mathematically sound, rational approach for determining when to stop sampling. Additionally, the metric naturally balances local refinement with global search in areas with high uncertainty. It is sometimes the case that a very inexpensive analysis in an area with very uncertain prior knowledge enhances the posterior beliefs about the space and provides more value than a very expensive analysis near the optimum, as shown in Figure 3.

To perform the actual VOI calculation in practice, the expected value of analyzing a point $x$ is defined as

$$E[u(\text{Analyze } X)] = \int_{-\infty}^{\infty} f_x(x) \max(E[u(X|x)], E[u(Y|x)]) dx$$

where $x$ is the point where the analysis is to be performed. $E[u(Y|x)]$ is the expected utility of the current best sampled point, and $E[u(X|x)]$ is the expected utility of choosing $X$ after analysis. Clearly only one choice will be made; if the utility of selecting $X$ is higher than that of selecting $Y$ (that is, the information from the analysis causes the designer to change her selection) then $X$ will be selected; otherwise, point $Y$, the current best achieved prior to the additional analysis, will be selected. The quantity $f_x(x)$ represents the distribution of possible outcomes and is computed as a Bayesian update of our prior beliefs about the truth at point $x$ and the posterior of the simulation results.

Since only one decision will be made to select $X$ or $Y$ (whichever has the higher utility), the expected utility of analyzing $x$ can be divided into a sum of two integrals as follows:

![Figure 3: Value of Information for Space Exploration](image-url)
where $T^*$ is the point at which the expected utility of selecting $X$ is equal to the expected utility of selecting $Y$; that is, the information provided by analyzing $x$ results in the decision maker being indifferent to selecting $X$ or $Y$. In this paper, a risk neutral utility function is assumed. In this case, 

$$E[u(x)] = \mu_x$$

which is the predicted mean from the surrogate model at point $x$, where we are considering additional analysis. The posterior distributions for $x$ given $X$ and for $y$ (which is the current best sampled point and independent of $X$) are then

$$X | x \sim N\left( \frac{\mu_x \sigma_x^2 + x \sigma_x^2}{\sigma_x^2 + \sigma_y^2}, \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \right)$$

$$y \sim N(\mu_y - C, \sigma_y)$$

where $C$ is the cost of the analysis. This cost is determined based on the time and computational resources required to run the analysis. The remaining terms are:

- $\sigma_x$: the square root of the MSE of the predictor at $x$
- $\sigma_y$: the standard deviation assigned to the model that would be used to analyze $x$
- $\mu_x$: the current best
- $\mu_y$: the standard deviation of the model that was used to determine the current best sampled point—this term actually falls out of the calculation in the case of risk neutrality

$T^*$ can be found as

$$T^* = \frac{(\sigma_x^2 + \sigma_y^2) \mu_x - \mu_x \sigma_x^2}{\sigma_x^2}$$

This leaves only $f_x(x)$, which is a probability density function characterizing the possible outcomes. To determine $f_x(x)$ for a particular analysis, we must ignore the observations from other models and consider only the observations from the model currently under consideration. It should be noted that VOI must be calculated separately for each model in the design space, and then the maximum of the maximum VOI’s from each model is taken to determine the next point and analysis. Therefore, if VOI for Model 1 is being calculated,

$$f_x(x) = \frac{1}{\sqrt{2\pi(\sigma_x^2)}} e^{-(x-\mu_s1)^2/2\sigma_x^2}$$

where $\mu_s1$ is the predicted mean at $x$ given the observations from model 1. This is calculated from the following predictor:

$$\mu_s1 = \hat{y}_s1 = \hat{y} + \hat{z}_s1 = \hat{y} + r^T y^*$$

where $\hat{y}$ is the prediction of the truth and $y^* = R_1^{-1}(Y_{s1} - \hat{y}(s))$

Here, $Y_{s1}$ are the observed samples from Model 1, $\hat{y}(s_1)$ is the prediction of the truth at the sample sites, and $R_1^{-1}$ corresponds to the correlation matrix for samples only from Model 1, that is $R_1 = R_1$ for $i = 1$.

The expected variance given the observations of Model 1, $\sigma_{y1}^2$, is determined from

$$\text{var}(\hat{y}_1) = \sigma_{y1}^2 = E[(\hat{y}_1 - \hat{y}_1)^2]$$

In this case, $\hat{y}_1 = y + z_1$ where $\hat{y}_1$ is the Gaussian process that perfectly models the behavior of Model 1, $y$ is the truth and $z_1$ is the Gaussian process that models the error between the two. Additionally, $\hat{y} = \hat{y} + \hat{z}_1$, which is the prediction of the mean given only the data from model one and the Gaussian process realizations used to predict the truth and the error between the truth and Model 1. Thus,

$$\sigma_{y1}^2 = E[(y - \hat{y})^2 + (z_1 - \hat{z}_1)^2]$$

The remainder of the calculation is rather lengthy, but the results for mean and variance with respect to Model 1 for the first iteration of the example presented in the next section are shown in Figure 4. The predicted mean passes through the design sites from Model 1. The variance is identically zero at the design sites because the model is deterministic; this is different from the estimate of the truth where there exists a non-zero variance at the design sites due to the recognition of model inadequacy.

**RESULTS AND DISCUSSION**

The algorithm, including the multi-accuracy kriging model and the Value of Information metric is implemented. As a preliminary test, numerical examples that satisfy the underlying assumptions are used as a case study. In this example, a truth model is generated using a Gaussian process so that the function satisfies the assumptions made in the initial model formulation; specifically, the truth can be modeled as a sum of a regression model and a Gaussian process realization. It is important to note that in practice the truth, or physical
function evaluations and very few high fidelity models.

In practice, the low and high fidelity models might correspond to a finite difference model and a differential equation based model for the same system. Additionally, the truth would be unknown. It is the job of the user, in this case, to assign an appropriate variance (\(\sigma^2_1, \sigma^2_2\)) to each model in order to characterize the models’ expected accuracy with respect to reality. In this example, the variance used to generate the model error for each model is also used to calculate the VOI. In practice, further research is required on how to characterize the models’ variance in a meaningful way. Generally speaking, the low fidelity models must have a higher variance than the high fidelity models in order for the higher fidelity observations to carry more ‘weight’ during the fitting of the surrogate model.

For this illustrative example, the initial surrogate model is seeded with five evenly distributed samples of the low fidelity model, as shown in Figure 5. This is simply an arbitrary choice of initial seeding; this could very well be done using a full factorial DOE or an LHS design, and this initial seeding will need to be considered in greater detail when multi-dimensional examples are considered.

For illustrative purposes, at each step of the optimization the VOI is calculated at the all points in the space for both the low and high fidelity analyses, as shown in Figure 6. The point and analysis combination with the highest VOI is then sampled in the next iteration and the surface is refit. This process continues until the maximum VOI is less than or equal to zero for all points in the design space.

Preliminary results of the prototype algorithm show a lot of promise. Often, the globally optimal solution is found with relatively few function evaluations and very few high fidelity function evaluations. Figures 5-12 show a sampling of steps from one run of the algorithm. In this particular run, the truth is generated using a Gaussian process with a \(\theta\) value of 15 and a \(\sigma^2\) of 20. The error between the truth and the low fidelity model is generated using a Gaussian process with \(\theta_1 = 1\) and \(\sigma^2_1 = 2\). The error between the truth and high fidelity model is generated using \(\theta_2 = 5\) and \(\sigma^2_2 = 0.01\). For each of the three models, 20 samples are randomly generated and a classic kriging model is fit to the data. For reproducibility, the kriging model parameters are given in Table 1. For this case a risk neutral preference is assumed, and the low and high fidelity analyses cost $0.01 and $1, respectively.

In this example, the optimizer converges to \(x = 0.5859, y = 8.0896\) after 7 low fidelity function evaluations (including the initial 5) and 1 high fidelity function evaluation. In the progression of model fits in Figures 5,7,9, and 11, it can be seen that the MSE of the predictor decreases as more design sites are added and the model becomes more accurate. Specifically, it is important to note that the MSE is reduced the most in the neighborhood of the optimum; this is a very desirable feature. It is important not to spend too many resources getting an accurate representation of the design space far from the optimum; rather, the uncertainty is reduced in those areas to ensure confidence in our solution, and the model is refined only in the neighborhood of the optimum. This is

<table>
<thead>
<tr>
<th>X</th>
<th>Gamma* Truth</th>
<th>Gamma* Low</th>
<th>Gamma* High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>4.885E+06</td>
<td>2.986E+04</td>
<td>-2.861E+01</td>
</tr>
<tr>
<td>0.105</td>
<td>-4.013E+07</td>
<td>-3.698E+05</td>
<td>2.218E+02</td>
</tr>
<tr>
<td>0.211</td>
<td>1.693E+08</td>
<td>1.789E+06</td>
<td>-8.714E+02</td>
</tr>
<tr>
<td>0.316</td>
<td>-4.872E+08</td>
<td>-4.477E+06</td>
<td>2.296E+03</td>
</tr>
<tr>
<td>0.421</td>
<td>1.071E+09</td>
<td>2.560E+06</td>
<td>-2.877E+06</td>
</tr>
<tr>
<td>0.526</td>
<td>-1.911E+09</td>
<td>-2.624E+06</td>
<td>6.931E+03</td>
</tr>
<tr>
<td>0.632</td>
<td>2.863E+09</td>
<td>2.723E+06</td>
<td>-8.452E+03</td>
</tr>
<tr>
<td>0.737</td>
<td>-3.683E+09</td>
<td>7.943E+03</td>
<td>-1.143E+04</td>
</tr>
<tr>
<td>0.842</td>
<td>4.117E+09</td>
<td>2.621E+06</td>
<td>-4.915E+03</td>
</tr>
<tr>
<td>0.947</td>
<td>-4.026E+09</td>
<td>-3.145E+06</td>
<td>-5.938E+01</td>
</tr>
<tr>
<td>1.053</td>
<td>3.443E+09</td>
<td>-3.523E+06</td>
<td>5.495E+03</td>
</tr>
<tr>
<td>1.158</td>
<td>-2.560E+09</td>
<td>5.872E+06</td>
<td>-9.669E+03</td>
</tr>
<tr>
<td>1.263</td>
<td>1.632E+09</td>
<td>1.273E+06</td>
<td>1.143E+04</td>
</tr>
<tr>
<td>1.368</td>
<td>-8.692E+08</td>
<td>-6.697E+06</td>
<td>-1.066E+04</td>
</tr>
<tr>
<td>1.474</td>
<td>3.674E+08</td>
<td>1.794E+06</td>
<td>8.154E+03</td>
</tr>
<tr>
<td>1.579</td>
<td>-1.095E+09</td>
<td>6.404E+06</td>
<td>-5.144E+03</td>
</tr>
<tr>
<td>1.684</td>
<td>1.380E+07</td>
<td>-8.403E+06</td>
<td>2.624E+03</td>
</tr>
<tr>
<td>1.789</td>
<td>5.563E+06</td>
<td>4.920E+06</td>
<td>-1.031E+03</td>
</tr>
<tr>
<td>1.895</td>
<td>-3.364E+06</td>
<td>-1.496E+06</td>
<td>2.795E+02</td>
</tr>
<tr>
<td>2.000</td>
<td>6.155E+05</td>
<td>1.934E+05</td>
<td>-3.914E+01</td>
</tr>
</tbody>
</table>

Table 1: Values for Generating Truth and Low and High Fidelity Data

For example, the truth might be the behavior of a car or other large system that is not yet prototyped, or even if the truth can be obtained using physical experiments, it may be too costly to do so. However, for illustrative purposes, a truth curve is included in this numerical example and we assume that it can be modeled as a correlated Gaussian process. From the truth, model inadequacy is captured using some different Gaussian processes to generate error functions. By summing these errors with the truth, low and high fidelity models of the truth are obtained. For this test case, random samples are drawn from a Gaussian process corresponding to the truth and then a classic kriging model is fit to the data. This process is repeated for the error between Model 1 (low fidelity) and the truth as well as the error between Model 2 (high fidelity) and the truth. These errors can then be added to the truth to generate the low and high fidelity models. It is these models from which observations are drawn in order to fit a Gaussian process surrogate surface as described in the previous section.
readily apparent in Figure 11 after a high fidelity sample is added right around the optimum.

The VOI is shown in Figures 6,8,10, and 12. It can be seen that the VOI tends to flatten out as more points are added, and the line at which the VOI converges to is equal to the cost of the analysis. This means that there is no anticipated gain by adding a sample in these regions, so all we stand to do is forfeit the cost of additional analysis. It is also noteworthy that the VOI for both models tends to be maximized in the same general areas. This is largely because the VOI calculation relies so heavily on a lot of the same data, regardless of the individual model cost and accuracy. When multi-dimensional examples are considered, it will be useful to see whether this is still the case and whether this information can be leveraged to improve efficiency in the process of maximizing the VOI. For this particular example, the final fit before the optimum is reached is shown in Figure 11, and the corresponding VOI, now less than or equal to zero for all points in the space, is shown in Figure 12.

Figure 5: Initial Seeding of Gaussian Process Model

Figure 6: VOI for each Analysis after Initial Seeding

Figure 7: Second Iteration of Gaussian Process Model

Figure 8: VOI for each Analysis after Second Iteration
CONCLUSIONS AND FUTURE WORK

In this paper, a Gaussian process based approach to surrogate modeling for multiple models is introduced along with a novel sampling criteria, VOI, which inherently accounts for costs incurred during the design process. This approach is illustrated using a numerical example problem. While the initial results are very promising, there are additional research opportunities to pursue.

There are still several algorithmic issues to be addressed. Sometimes a misleading fit is encountered, often due to undersampling. This can potentially be avoided using a more intelligent initial sample set or by implementing a cross validation scheme. Another option may be to tune the risk aversion parameter to slightly more risk averse, thereby assuring more global exploration. Aside from the misleading fits, numerical conditioning can cause poor predictions in cases where there are too many samples in a particular area; this clustering of design sites produces a near singularity in the covariance matrix making inversion numerically unstable. The ill-conditioning problems can be addressed from a couple of perspectives, either by modifying the calculation of the inverse of the covariance matrices to make them more numerically tractable or by scaling and spacing out the samples. Additional algorithmic efficiency improvements may also be made in the maximizations of likelihood and of value of information.

Beyond the algorithm itself, this paper provides only an illustration of the method; more rigorous testing and characterization is needed. Comparisons will be made to other similar approaches, and the intent is to define a test suite for consistent comparison across methods. Also, more rigorous engineering examples will need to be explored with this new approach.
ACKNOWLEDGMENTS
This work has been funded by the ERC for Compact and Efficient Fluid Power, by the National Science Foundation under Grant No. EEC-0540834.

REFERENCES


