Eng. Opt., 2002, Vol. 34, pp. 263–278

Taylor & Francis
Taylor & Francis Group

EXPLORATION OF METAMODELING SAMPLING CRITERIA FOR CONSTRAINED GLOBAL OPTIMIZATION

MICHAEL J. SASENA[a,*], PANOS PAPALAMBROS[a,†] and PIERRE GOOVAERTS[b,‡]

[a]*Mechanical Engineering Department 3200, EECS Building, University of Michigan, Ann Arbor, MI 48109-2125, USA,* [b]*Civil and Environmental Engineering Department, 117 EWRE Building, University of Michigan, Ann Arbor, MI 48109-2125, USA*

(Received 11 December 2000; In final form 23 August 2001)

The use of surrogate models or metamodeling has lead to new areas of research in simulation-based design optimization. Metamodeling approaches have advantages over traditional techniques when dealing with the noisy responses and/or high computational cost characteristic of many computer simulations. This paper focuses on a particular algorithm, Efficient Global Optimization (EGO) that uses kriging metamodels. Several infill sampling criteria are reviewed, namely criteria for selecting design points at which the true functions are evaluated. The infill sampling criterion has a strong influence on how efficiently and accurately EGO locates the optimum. Variance-reducing criteria substantially reduce the RMS error of the resulting metamodels, while other criteria influence how locally or globally EGO searches. Criteria that place more emphasis on global searching require more iterations to locate optima and do so less accurately than criteria emphasizing local search.

Keywords: Approximation; Computer simulation; Global optimization; Kriging

1 INTRODUCTION

When simulations become computationally expensive, the number of simulation-based function evaluations required for optimization must be carefully controlled. To that end, researchers have explored the use of metamodels, namely, simpler approximate models calibrated to sample runs of the original simulation. The approximate model or metamodel can replace the original one, thus reducing the computational burden of evaluating numerous designs.

One particular algorithm used in this way is the *Efficient Global Optimization* (EGO) algorithm developed by Schonlau, Welch and Jones [9]. EGO starts with a small data sample within the design space and fits a metamodel via kriging [7]. Based on this metamodel, a set of additional points, the so-called infill samples, are selected to be evaluated by the full simulation. The sample set is then updated, the model refit, and the process of choosing new points continues until the improvement expected from sampling additional points has

* Corresponding author. E-mail: msasena@umich.edu
† E-mail: pyp@umich.edu
‡ E-mail: goovaert@umich.edu

ISSN 0305-215X print; ISSN 1029-0273 online
DOI: 10.1080/03052150290030389

become sufficiently small. In this study only one infill sample at a time is added, but the algorithm is not restricted to do so.

Different criteria have been developed for selecting infill sample points. EGO uses a *generalized expected improvement* function whereby points that have either low objective function value or high uncertainty (*i.e.* model inaccuracy) are preferred. A single parameter, g, determines the balance between the two trends. Researchers in the field of geostatistics have proposed other criteria for studies involving the sampling of contaminated sites. No studies to date have shown how these criteria may behave within an optimization algorithm such as EGO. Here, the influence of the parameter g and the choice of sampling criterion are considered. An early discussion of the issues examined here was presented in Sasena *et al.* [12].

The article is organized as follows. First, the expected improvement function is described based on Schonlau's dissertation [13]. Next, three criteria proposed by Watson and Barnes [14] and the maximum variance criterion are reviewed. Four analytical examples are then presented to demonstrate the differences between the criteria. An examination of how to handle constraints ensues. Finally, a simulation-based example is presented, and some general conclusions are drawn.

2 GENERALIZED EXPECTED IMPROVEMENT

Let $f^n_{\min}$ be the minimum feasible sampled value of the function $y = f(\mathbf{x})$ after n evaluations, where $\mathbf{x}$ is a vector of input values, The response is treated as a realization of a random variable $Y(\mathbf{x})$ assumed to be Gaussian with variance $s^2(\mathbf{x})$. For simplicity, the dependence on $\mathbf{x}$ is left out, denoting the values $y(\mathbf{x})$ as y and $s(\mathbf{x})$ as s. The improvement over the current best point is defined as

$$I = \max\{0, f^n_{\min} - Y\} \tag{1}$$

Using kriging to predict $\hat{y}$ and $\hat{s}$, the expected improvement is computed as

$$E(I) = \begin{cases} (f^n_{\min} - \hat{y})\Phi\left(\dfrac{f^n_{\min} - \hat{y}}{\hat{s}}\right) + \hat{s}\phi\left(\dfrac{f^n_{\min} - \hat{y}}{\hat{s}}\right) & \text{if } \hat{s} > 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \tag{2}$$

where $\Phi(\ )$ and $\phi(\ )$ denote the cumulative distribution function (cdf) and the probability density function (pdf) of the standard normal distribution, respectively.

Inspection of the expected improvement function reveals two important trends. The first term in Eq. (2) is the difference between the current minimum and the predicted value multiplied by the probability that $Y(\mathbf{x})$ is smaller than $f^n_{\min}$ and is therefore large where $\hat{y}$ is likely smaller than $f^n_{\min}$. The second term is the standard deviation of $y(\mathbf{x})$ multiplied by the probability that $y(\mathbf{x})$ is equal to $f^n_{\min}$. This term is large where there is high uncertainty in the value of the prediction itself, hence whether the current minimum will be lowered or not. As the prediction variance vanishes at the sampled data points, the expected improvement also vanishes.

The generalized form of the expected improvement introduces a non-negative integer parameter g, as

$$I^g = \max\{0, (f^n_{\min} - Y)^g\} \tag{3}$$

The resulting recursive formula for the generalized expected improvement is

$$E(I^g) = s^g \sum_{k=0}^{g} (-1)^k \left(\frac{g!}{k!(g-k)!} \right) (f'^n_{\min})^{g-k} T_k \tag{4}$$

where $f'^n_{\min} = (f^n_{\min} - \hat{y})/\hat{s}$ and

$$T_k = -\phi(f'^n_{\min})(f'^n_{\min})^{k-1} + (k-1)T_{k-2} \tag{5}$$

starting with $T_0 = \Phi(f'^n_{\min})$ and $T_1 = -\phi(f'^n_{\min})$.

The impact of the parameter g in Eq. (4) is illustrated using a one-dimensional example in Figure 1. The w-shaped dashed line is the true objective function we wish to model, while the solid line is the kriging approximation conditional to the sample points shown as circles. The function at the bottom is the expected improvement (EI) function, normalized to make better comparisons across infill criteria. For $g = 1$ (left graph), EI rises significantly in two areas: the region on the left is sparsely sampled, leading to high model uncertainty, while the region on the right is where the expectation of finding a better objective function value is high. For $g = 5$ (right graph), the EI function rises only in the left region where model uncertainty is high. In summary, increasing the value of g shifts emphasis towards global search by giving more importance to the second term in Eq. (2).

Although the value used for g is extremely significant, there is no obvious way to select it. Too high a value could prevent EGO from converging on a good solution in a reasonable amount of time. Too low a value could lead EGO to overlook areas of high uncertainty as it searches too locally. By analogy to the simulated annealing algorithm, which starts searching globally then refines the search more locally as iterations continue, we propose to start with a large g value, reducing the value towards zero. The heuristic cooling schedule used in this work is shown in Table I. The use of the EI function with this schedule is referred to as the *Cool* criterion in this article.

3 ALTERNATIVE SAMPLING CRITERIA

Watson and Barnes describe three criteria to select a set of locations for further sampling once an initial set of data has been collected [14]. Each criterion aims to solve a problem with a different objective, namely, to (i) locate the threshold-bounded extreme, (ii) locate

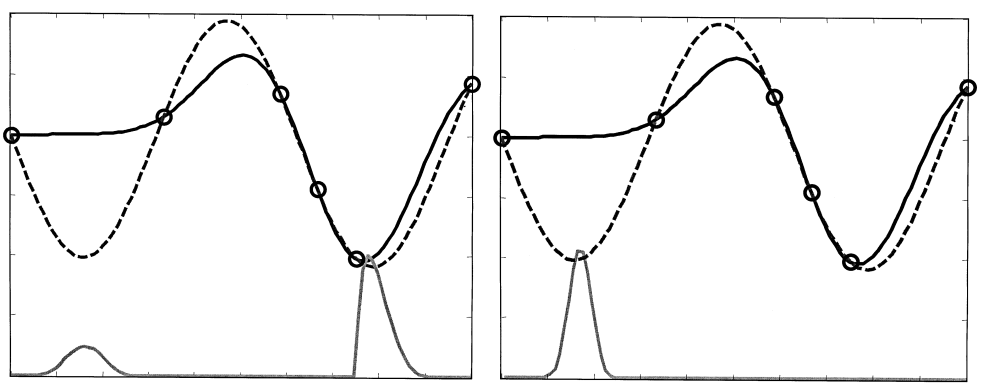

FIGURE 1 EI function for $g = 1$ (left) and $g = 5$ (right).

TABLE I Cooling Schedule.

Iteration	*g value*
1–4	20
5–9	10
10–19	5
20–24	2
25–34	1
≥35	0

the regional extreme, or (iii) minimize surprises. They are abbreviated here as WB1, WB2, and WB3, respectively. One additional criterion, the maximum variance, is described as well. Plots for each criterion in Figure 2 are normalized to facilitate comparisons.

3.1 Locating Threshold-Bounded Extremes

This criterion has been developed in the context of contamination testing, and the objective was to locate points that maximize the probability that at least one of the infill samples exceeds some specified threshold. Here this concept is extended to design optimization by

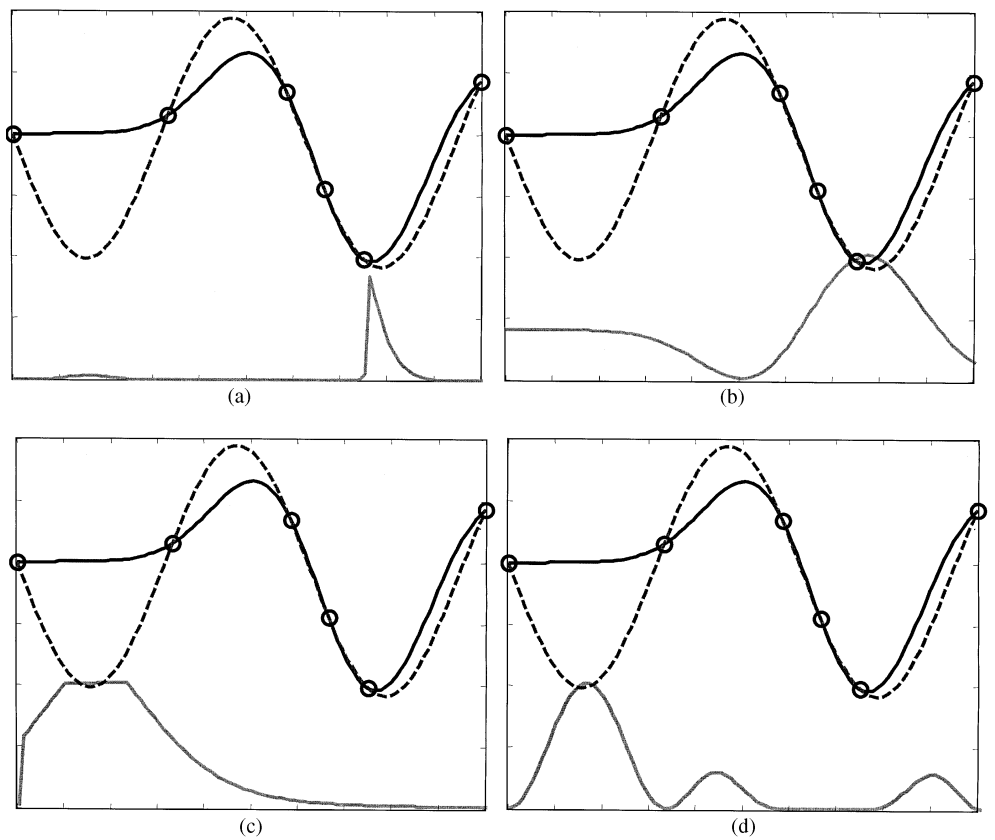

FIGURE 2 Criteria WB1 (a), WB2 (b), WB3 (c), and Maxvar (d) each normalized to facilitate comparisons.

using $f^n_{\min}$ as the threshold. The objective is now to maximize the probability of being no greater than $f^n_{\min}$, which is computed as

$$\text{WB1} = \Phi\left(\frac{f^n_{\min} - \hat{y}}{\hat{s}}\right) \tag{6}$$

Notice that this formulation is in fact the EI function with $g = 0$ and is thus extremely local in its search. One must therefore be confident that the model has found the region of the optimum for this criterion to be successful.

The behavior of this function is illustrated in Figure 2(a). The large peak on the right is due to the fact that the cdf of a normal distribution is strictly increasing, making WB1 largest for positive quantities, *i.e.* when the predicted value is smaller than the current best point. The smaller peak in the undersampled region on the left is due to the large uncertainty, where the large $\hat{s}$ lowers the magnitude of the negative argument of the cdf, thereby increasing the value of WB1.

3.2 Locating the Regional Extreme

The second criterion attempts to minimize the expected value of the smallest observation once the infill samples have been added:

$$\text{WB2} = \begin{cases} \hat{y} + (f^n_{\min} - \hat{y})\Phi(f'^n_{\min}) + \hat{s}\phi(f'^n_{\min}) & \text{if } \hat{s} > 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \tag{7}$$

The resulting formula is remarkably similar to the EI function in Eq. (2). The only difference is the additional first term, which is the predicted value at the location of interest. This criterion thus gives slightly more merit to local search than does the EI function. It is also smoother since it does not return to zero at the sampled points, see Figure 2(b). This appealing trait is noteworthy, as it may help in locating the maximum of the criterion.

3.3 Minimizing Surprises

The third criterion aims to minimize the maximum probability that a true value deviates significantly from its predicted value. Watson and Barnes use the simplified expression

$$\text{WB3} = \min_{\mathbf{x}} \max_{\mathbf{v}} \{\text{Var}[Y(\mathbf{v})|\mathbf{S} \text{ and } \mathbf{x}]\}, \tag{8}$$

where $\mathbf{x}$ is the candidate infill sample point of interest, $\mathbf{v}$ is a generic location in the design space, and $Y(\mathbf{v})$ is the random variable at the unobserved location $\mathbf{v}$. Notice that the variance is conditional to both the sample set, $\mathbf{S}$, and the candidate infill samples, $\mathbf{x}$. One can compute the variance of $Y(\mathbf{v})$ because the updated variance (*i.e.* the variance of the model once the candidate infill samples have been added) does not depend on sampled values, but is a function of only their locations and a given covariance function. WB3 requires locating the maximum variance for any given candidate infill sample location. This minimax problem within the original design optimization problem adds significantly to the total run time. Thus it is best suited for problems where the objective function is extremely expensive to calculate.

A characteristic of the updated variance of $Y(\mathbf{v})$ is that it becomes constant if the distance between $\mathbf{v}$ and the closest sample point exceeds the range of correlation [7]. As a consequence, the profile of WB3 values in Figure 2(c) is flat on the left side, which is under-

sampled. Also, adding infill samples far from areas that are not "covered" will not reduce the maximum variance. These characteristics can cause serious difficulties in locating the maximum of WB3.

3.4 Maximum Variance

Because of the inherent difficulty in solving the minimax problem of WB3, a simpler measure of uncertainty is needed. The *Maxvar* criterion uses the variance at the candidate location and is to be maximized. While WB3 has the benefit of looking ahead to locate regions of high uncertainty in the *next* iteration, Maxvar is much more reliable and easy to compute, see Figure 2(d).

4 ANALYTICAL EXAMPLES

The impact of each of the above criteria on EGO is assessed using four analytical examples. Test problems with a small number of variables were chosen for better visualization. Readers are directed to Schonlau [13] to see how EGO handles problems of six to ten design variables. Problems with a large number of variables may be solved using a decomposition strategy [4].

Example 1 (One-dimensional function) The first example is a one-dimensional function with two local minima (see Fig. 3(a)) and is defined as

$$f = -\sin(x) - \exp\left(\frac{x}{100}\right) + 10 \tag{9}$$

with $x \in [0, 10]$. A local minimum of 7.9841 occurs at $x = 1.5810$ (circle), while the global minimum of 7.9182 is at $x = 7.8648$ (asterisk).

Example 2 (Branin function) The second example is the Branin test function [3], defined as

$$f = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10 \tag{10}$$

with $x_1 \in [-5, 10]$ and $x_2 \in [0, 15]$. The three global minima at $\mathbf{x} = \{3.1416,\ 2.2750\}$, $\mathbf{x} = \{9.4248, 2.4750\}$ and $\mathbf{x} = \{-3.1416, 12.2750\}$ shown as asterisks in Figure 3(b) have identical function values of 0.3979.

Example 3 (Two-dimensional function) A sinusoidal constraint is placed on a multimodal function in two dimensions for the third example.

$$\begin{aligned} &\text{minimize } f = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2) \\ &\text{subject to: } -\sin(x_1 - x_2 - \pi/8) \le 0 \end{aligned} \tag{11}$$

The objective function is defined over the range $x_i \in [0, 5]$, $i = 1, 2$, and the constraint is active at the true solution. The global solution has a value of -1.1743 at $\mathbf{x} = \{2.7450, 2.3523\}$. Figure 3(c) shows the contour plot of the objective function and constraint boundary (the diagonal lines). The hash marks indicate the infeasible side of the constraint. The unconstrained and constrained minima are depicted by a circle and an asterisk, respectively.

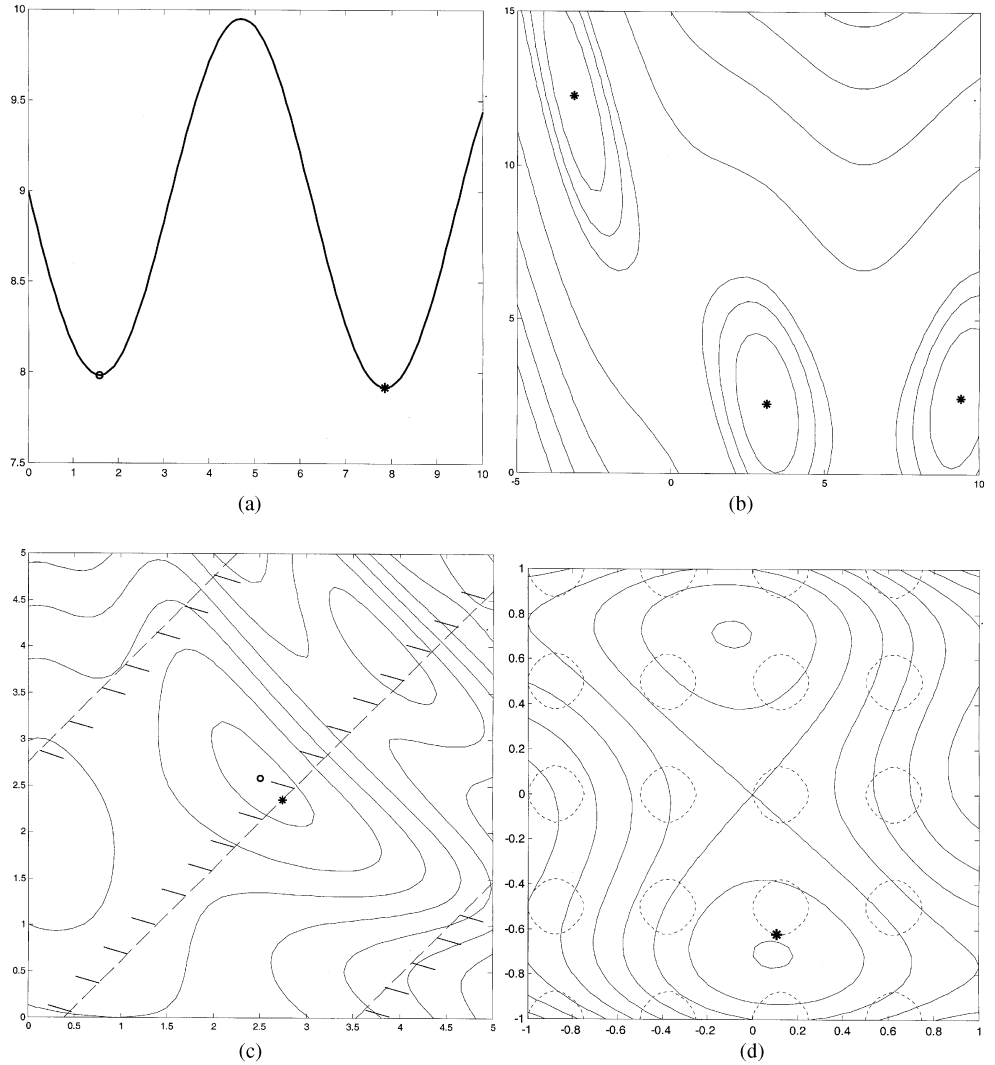

FIGURE 3 Analytical test problems. Example 1: one-dimensional function (a), Example 2: Branin function (b), Example 3: two-dimensional function (c), Example 4: Gomez #3 function (d).

Example 4 (Gomez #3 function) The Gomez #3 test function [6] is a difficult example because the constraint cuts the feasible design space into several small islands. The problem is

$$\begin{aligned} &\text{minimize } f = \left(4 - 2.1x_1^2 + \frac{1}{3}x_1^4\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 \\ &\text{subject to: } -\sin(4\pi x_1) + 2\sin^2(2\pi x_2) \le 0 \end{aligned} \tag{12}$$

In Figure 3(d), the objective function is shown in solid lines, and the islands of feasible design space lie inside the dashed circles. The global optimum lies at $\mathbf{x} = \{0.1093, -0.6234\}$ with a value of -0.9711, depicted by an asterisk.

4.1 Methodology and Comparison Metrics

In order to isolate the impact of the infill criteria for the analytical examples it was necessary to eliminate the influence of model accuracy. Thus for each function of Figure 3, the same kriging model parameters were used for each criterion and for all iterations. Typically, the kriging model parameters are fit at each iteration via Maximum Likelihood Estimation (MLE) [13]. In preliminary testing, it was observed that cross-validation fitting often provides more accurate kriging models than MLE fitting [12]. Thus a set of observations collected from previous experimentation was used to minimize the Root Mean Square error of cross-validation. While cross-validation was more expensive, no further model fitting was required during optimization. The reduced overhead more than made up for heavy computational costs incurred early in the design process.

In this work, the optimum of the infill sampling criteria was found via the DIRECT algorithm due to Jones [8]. It does not use gradient information and has global search properties that help it avoid the local optima inherent to the infill criteria. For the examples here, 30 iterations of DIRECT were used, requiring on the order of 300 evaluations of the infill criteria.

Comparisons of the infill sampling criteria are difficult because there is no rigorous convergence criterion for EGO. Schonlau proposed stopping the search once the ratio of the expected improvement to the current best sample value becomes sufficiently small [13]. Because this rule has no meaning for the alternative criteria studied here, each test is stopped after 100 function calls have been made, and the following metrics are computed:

- $x_{1\%}$ *metric*: The number of function calls required before a feasible point is sampled within a box the size of $\pm 1\%$ of the design space range centered around the true solution.
- $f_{1\%}$ *metric*: The number of function calls required before a feasible point is sampled with a value within 1% of the true solution.
- x_* *metric*: The Euclidean distance from the best feasible sample point to the global solution, x_*.
- *RMS metric*: The global modeling error. After 100 evaluations, the metamodel is compared to the true function on a 50 by 50 gridded set of locations (100 locations in the case of Example 1). The resulting errors at $N = 2500$ or 100 points are summarized by the RMS error, calculated as

$$RMS = \frac{1}{N}\left(\sum_{i=1}^{N}(\hat{y}(x_i) - f(x_i))^2\right)^{1/2} \tag{13}$$

The first two metrics measure how efficiently the algorithm finds the solution, while the third one measures how accurately EGO finds the solution. The last metric evaluates how accurately the final metamodel approximates the design space. To facilitate comparison, the last two metrics have been normalized as percentages of the largest values in each example. For all metrics, lower values are better.

4.2 Analytical Results

The bar graph in Figure 4(a) shows the number of function calls required by each criterion to reach $x_{1\%}$ for the four examples. In the legend, the number following EI (*e.g.* EI5) denotes the constant value of g used in Eq. (4). Similarly, the *Cool* criterion refers to the EI criterion using g values as described by the cooling schedule of Table I. Variance-reducing techniques (WB3 and Maxvar) by far are the worst at locating the optimum. Since they do not aim at locating even local minima, there is no reason to expect they will sample a point within

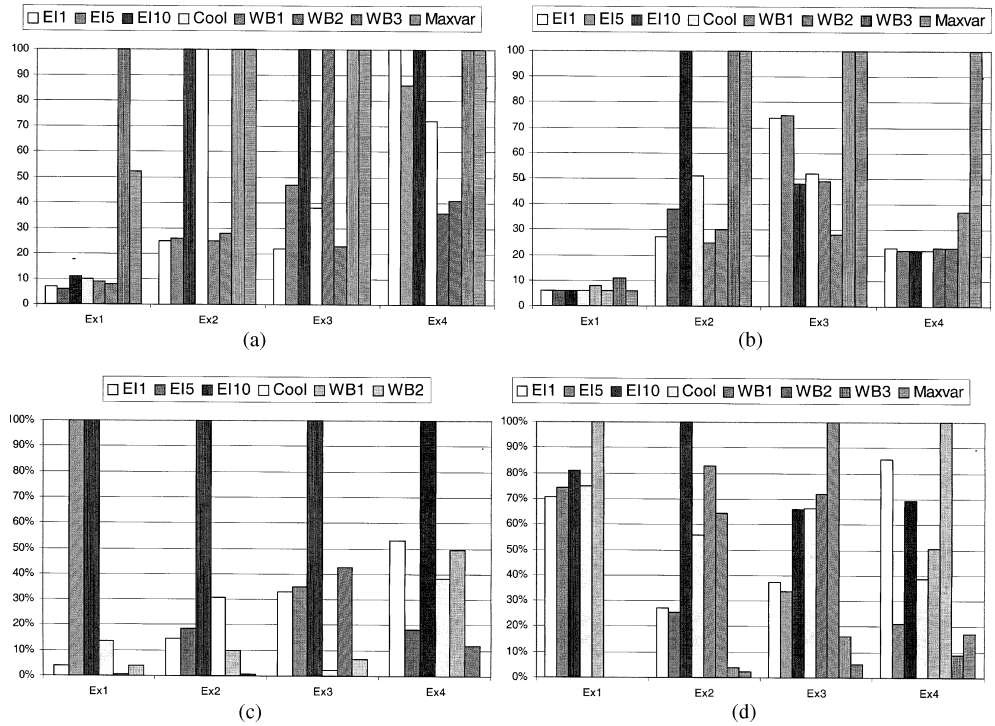

FIGURE 4 Analytical example results for the $x_{1\%}$ metric (a), $f_{1\%}$ metric (b), x_* metric (c) and RMS metric (d).

the $x_{1\%}$ limits, except by chance. The next worst criterion is EI10, the most globally searching one of the remaining criteria. Clearly, criteria that try to balance improving model accuracy with locating minima perform much better at sampling points near the optimum. Similar results are found for the $f_{1\%}$ metric (see Fig. 4(b)). Again, the variance-reducing criteria take the longest to reach the goal, if they do so at all. The other criteria, however, show no consistent trends from one example to another, which precludes general conclusions.

Because the variance-reducing criteria seldom found the $x_{1\%}$ region, the x_* metric was computed for only the first six criteria (see Fig. 4(c)). The most globally searching method, EI10, stays the furthest from the optimum in all examples. The more locally searching criteria (EI1, WB2 and cool) tend to perform better. While the most locally searching method, WB1, most accurately locates the optimum for the first two examples, that is not the case for the constrained examples. This may be due to the method for handling constraints which will be further investigated in the next section. Notice that the local accuracy does not systematically deteriorate as the value of g in the EI function increases. This unexpected result could originate from the inability to determine practically the exact optimum of the infill criterion for every iteration.

Figure 4(d) shows that the variance-reducing criteria consistently yield the lowest RMS, sometimes orders of magnitude smaller than the RMS obtained using other criteria. Despite the relative success of the variance-reducing criteria at spreading data points, some unexpected outcomes were observed. For example, since by design EI10 searches more globally than the EI criteria with lower g values, its RMS metric should be smaller. This was not the case and warrants further investigation. An additional question focuses on the WB3 criterion. Figure 5 shows the contour of the true Branin function with circles indicating the initial 21-point design and the x's marking the remaining 79 infill sample points. Notice the clustering of samples in

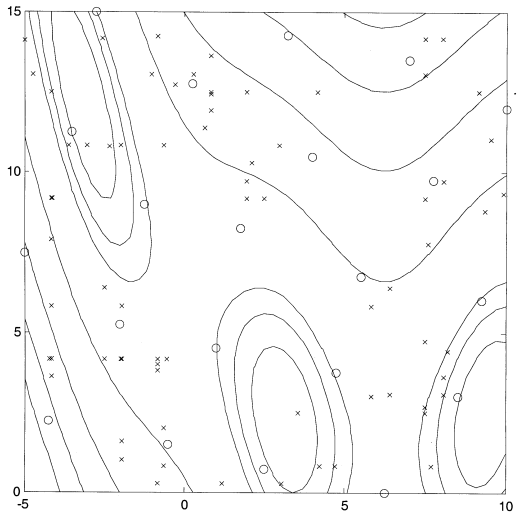

FIGURE 5 Branin example using WB3.

several locations. If the criterion had successfully been implemented, one would have expected the infill sample points to spread out more evenly. One possible explanation is the failure of the internal optimization algorithm to find a good solution to the difficult minimax problem. Searching WB3 more thoroughly at each iteration could produce better results but is not practical because the overhead associated with the problem is already staggering.

5 CONSTRAINT HANDLING

One major concern with the EGO algorithm is the handling of constraints. The current literature tentatively suggests multiplying the value of the expected improvement by the probability that the point is feasible [13]. However, the value of the infill criterion may be impacted too strongly, keeping the algorithm from exploring points directly along the constraint boundary where the true optimum lies. A remedy proposed here is to use a penalty method, whereby a large constant (*i.e.* a penalty) is added to the criterion in order to restrict it from choosing infill samples in the infeasible region.

Figure 6(a) shows the region near the optimum for Example 3. The contours of the kriging approximation of the objective function are shown as dashed lines. The constraint boundary cuts diagonally across, the bottom right being the feasible side. The circle on the top left is the nearest sample point, and the asterisk near the center is the location of the true optimum. The fact that the true optimum lies directly on the bound of the constraint metamodel is an indication that the constraint was approximated accurately. The solid contour lines are the EI criterion assuming there is no constraint. The optimum of the criterion is shown as a triangle.

Figure 6(b) shows the contours of the probability-adjusted EI criterion and Figure 6(c) the contours of the penalty-adjusted EI criterion. The optimum for each criterion is shown as a triangle. Although both criteria drop off quickly in the infeasible region, the optimum of the penalty method is much closer to the constraint boundary than the probability method. This provides support to the claim that the probability method effectively pushes the infill sample point away from the constraint.

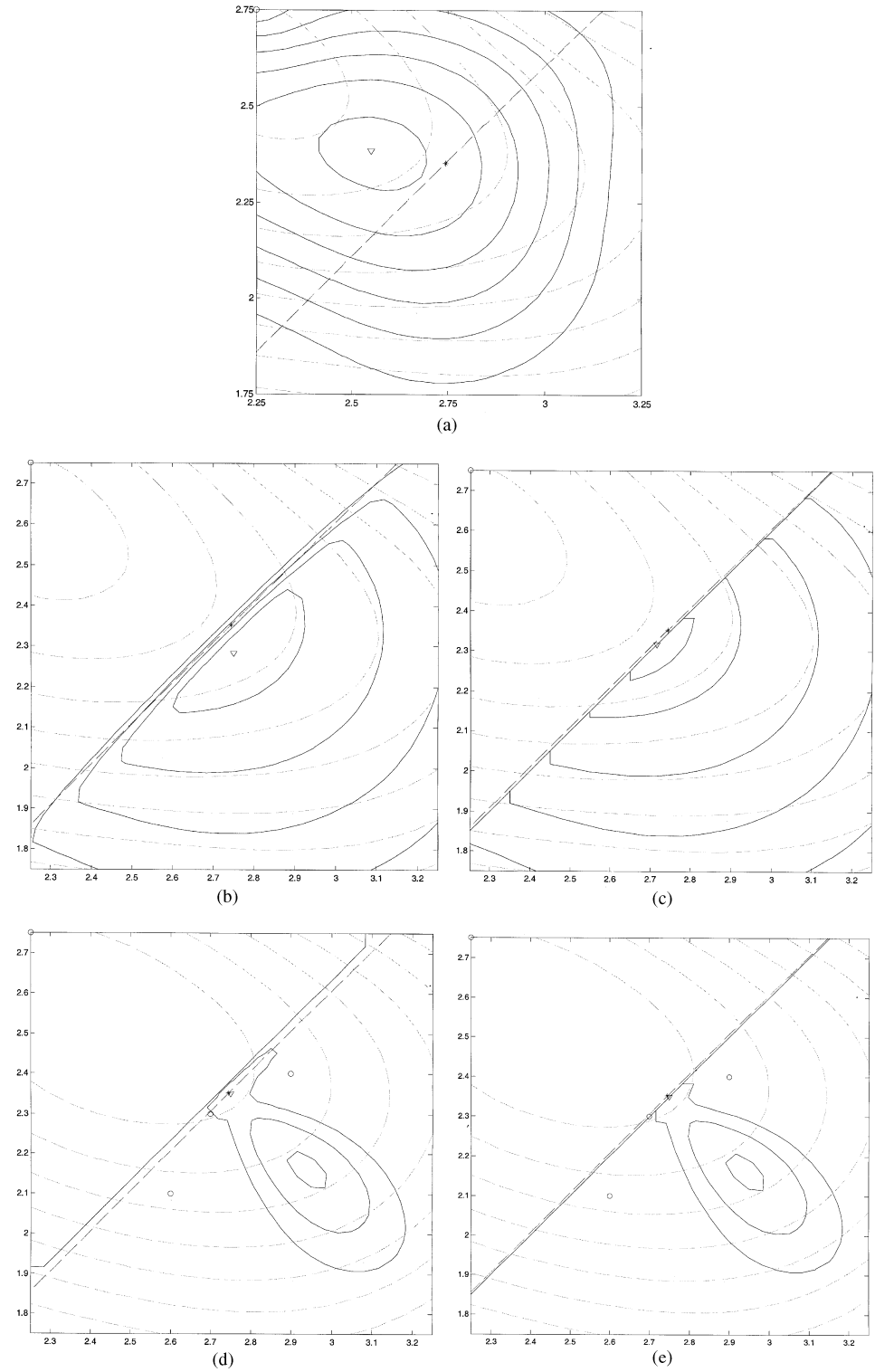

FIGURE 6 EI contours for Example 3. Unconstrained EI (a), the probability-adjusted EI (b) and penalty-adjusted EI (c). The impact of additional samples on probability-adjusted EI (d) and penalty-adjusted EI (e).

A similar test was run adding three sample points around the optimum. Figures 6(d) and 6(e) show the contours of the probability- and penalty-adjusted criteria, respectively, with the additional sample points shown as circles and the optimum infill points shown as triangles. The two are extremely similar in shape and in the location of their optima. The probability method drops off slightly past the constraint boundary whereas the penalty method drops exactly at the boundary. The implications are unclear. Further testing must be done to understand what happens as points pile up near the constraint boundary or as the number of constraint functions increases.

To visualize how the different methods behave on a full optimization problem, Example 3 was run for 100 function evaluations. Figure 7 shows a comparison of the function calls made by the EI criterion with the probability method throughout (left) and with a switch to the penalty method after ten iterations (right). The contours are of the true functions, and the initial

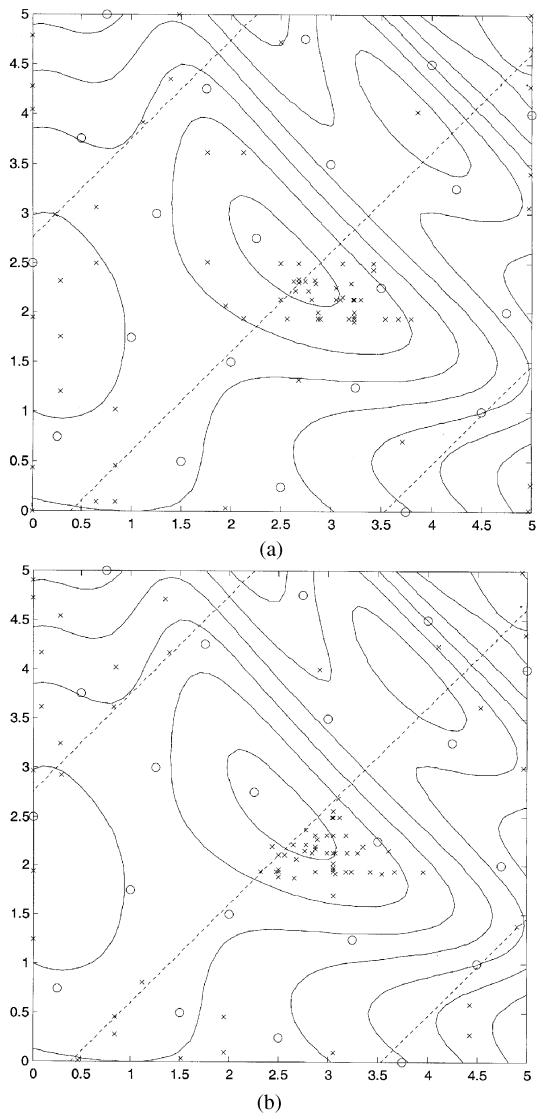

FIGURE 7 Differences in the feasibility of sampling regions for the probability (a) and penalty (b) methods.

samples are shown as circles and the infill points as ×'s. While both methods cluster around the global optimum, the probability method samples more frequently in the infeasible space. The penalty method clusters more sample points around the optimum on the feasible side of the constraint. For optimization problems where strict feasibility is important, using a penalty method may prove useful.

Tests were run on Example 3 with most of the criteria to compare the probability and penalty approaches. The variance-reducing criteria were left out because the presence of constraints does not impact the measure of model uncertainty. Figure 8 (left graph) shows a comparison of the probability (*e.g.* EI1) and penalty (*e.g.* EI1pen) methods for the $x_{1\%}$ and $f_{1\%}$ metrics. The penalty method was able to find a point in the $x_{1\%}$ region using fewer or the same number of function evaluations than the probability method for every criterion except EI5. As for the $f_{1\%}$ metric, the penalty method actually took the same or *more* function evaluations, which may reflect the fact that the $f_{1\%}$ goal is more difficult to satisfy than the $x_{1\%}$ goal in this example. Thus, the penalty method, which relies more heavily on the accuracy of the constraint model, takes longer to refine the solution, even though it located the $x_{1\%}$ area more quickly.

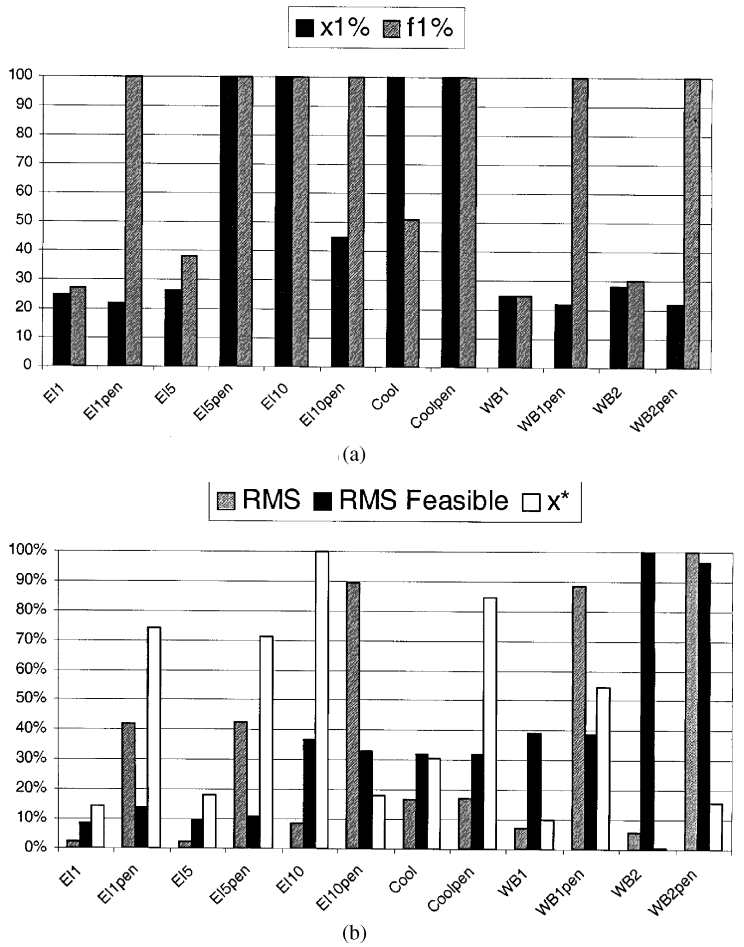

FIGURE 8 Comparison metrics for penalty and probability adjusted criteria for Example 3 – $x_{1\%}$ and $f_{1\%}$ (a) and x_* and RMS for either the entire design space or just the feasible portion thereof (b).

For all criteria, the probability method resulted in a lower RMS error, see Figure 8 (right graph). This is expected since the penalty method keeps points away from the infeasible region as illustrated in Figure 7. The computation of the RMS errors in just the feasible region of the design space indicates that the penalty method does not always produce higher RMS errors, meaning the models produced by the penalty method are not necessarily less accurate in the feasible design region. Contrary to expectations, the penalty method provided lower accuracy solutions for all but the EI10 criterion.

6 SIMULATION-BASED EXAMPLE

In this section, a hybrid electric vehicle (HEV) simulation called ADVISOR [10] is used to explore the capabilities of EGO to work with computer simulations. On an Ultra 10 SunSparc workstation, ADVISOR requires approximately two minutes to compute the fuel economy and acceleration performance of a given design. More details on the simulation can be found in Refs. [1,5].

The design problem illustrated here is to maximize the fuel economy in miles per gallon (m.p.g.) of a mid-sized hybrid electric passenger car subject to a set of performance constraints, g_1 to g_8. The first two constraints pertain to the maximum speed and acceleration the vehicle can obtain. The third requires that five seconds into an acceleration test, the vehicle must travel at least 140 feet. Constraints four though six limit the time required for the vehicle to accelerate to a given speed. Constraint seven ensures that the vehicle can sustain 55 mph speeds on at least a 6.5% grade slope, while the last constraint ensures it can start forward movement from a dead stop on at least a 30% grade slope. The design variables are the size of the engine, the size of the electric motor, and the size of the battery pack. The problem formulation is summarized below.

$$
\begin{aligned}
&\text{maximize} \quad f(\boldsymbol{x}) = \text{m.p.g.} \\
&\quad \boldsymbol{x} = \{\text{engine size, motor size, battery size} \\
&\text{subject to:} \\
&\quad g_1 : \text{maximum speed} \geq 85\,\text{mph} \\
&\quad g_2 : \text{maximum acceleration} \geq 0.5\,\text{g's} \\
&\quad g_3 : 5 \text{ second distance} \geq 140\,\text{feet} \\
&\quad g_4 : 0\text{–}60\,\text{mph time} \leq 12 \text{ seconds} \\
&\quad g_5 : 0\text{–}85\,\text{mph time} \leq 23.4 \text{ seconds} \\
&\quad g_6 : 40\text{–}60\,\text{mph(passing)time} \leq 5.3 \text{ seconds} \\
&\quad g_7 : 55\,\text{mph (cruising) gradability} \geq 6.5\% \\
&\quad g_8 : \text{maximum launch grade} \geq 30\% \\
&\text{and} \\
&\quad 15\,\text{kW} \leq \text{engine size} \leq 150\,\text{kW} \\
&\quad 5\,\text{kW} \leq \text{motor size} \leq 50\,\text{kW} \\
&\quad 5 \text{ modules} \leq \text{battery size} \leq 70 \text{ modules}
\end{aligned}
\tag{14}
$$

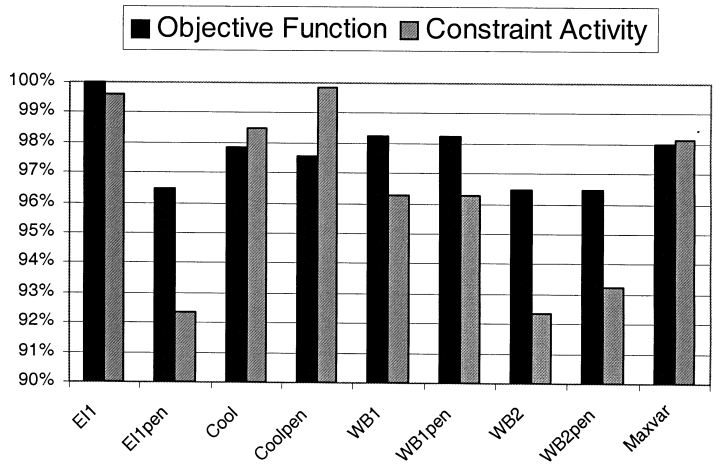

FIGURE 9 Simulation example results.

Nine tests were run on the simulation-based example using 200 function evaluations in each case. Because of the poor performance of the WB3 criterion in locating minima and its extreme overhead, the Maxvar criterion is the sole variance-reducing method presented. Both the probability and penalty methods were applied to the remaining criteria.

Results show the 40–60 passing time constraint as closest to being active in all cases which agrees with prior experience with the HEV simulation. The infill criteria were compared by looking at the values of the objective function and the passing time constraint at each test's best feasible sample point. All the fuel economy numbers were then normalized by the best design observed (EI1 in this case). The 40–60 time numbers were also normalized by dividing the actual time by the goal of 5.3 seconds. The closer to unity, the closer the design came to a solution with an active constraint. Thus, for both bars in Figure 9, results closer to 100% are better.

The penalty method once again did not perform as hoped. In each case, the probability method found feasible designs that were no worse than the penalty method and significantly better in the case of the EI criteria. As for the constraint activity, the penalty method obtained results closer to the constraint boundary for just two of the four criteria, The EI criteria performed the best in this case study, providing both the best fuel economy and a nearly active constraint. Other criteria do not exhibit clear trends. The *Coolpen* criterion (*i.e.* the *Cool* criteria using the penalty method for the constraints) most accurately located a constraint boundary, but the resulting design had worse fuel economy. The most locally searching criterion, WB1, found the second best design even though it had quite a bit of slackness in the performance constraint.

7 CONCLUSION

The work presented increases the flexibility of a global metamodeling optimization algorithm by proposing additional infill sampling criteria. While no single criterion is best in all senses, the framework allows the designer to change the criterion at each iteration in order to take advantage of potential benefits as documented by the present results. For example, the first few iterations may be spent increasing the global model accuracy via the Maxvar criterion.

The next few iterations could then refine current locations of promise with more locally searching criteria, such as EI or WB2. One could alternate between these criteria before refining the search with a few iterations using WB1.

Another important issue raised is how to handle constraints. Preliminary tests do not show an advantage for the penalty method over the probability method at locating bound-constrained optima. One problem is the inherent limitation of transforming a constrained objective into an unconstrained criterion using the probability, penalty, or similar methods. Better ways to handle the constraints may require completely new approaches such as those suggested by Sasena *et al.* [11] or Audet *et al.* [2].

Acknowledgements

This research was partially supported by the General Motors Collaborative Research Laboratory and the Automotive Research Center at the University of Michigan. This support is gratefully acknowledged.

References

[1] Assanis, D., Delagrammatikas, G., Fellini, R., Filipi, Z., Liedtke, J., Michelena, N., Papalambros, P., Reyes, D., Rosenbaum, D., Sales, A. and Sasena, M. (1999). An optimization approach to hybrid electric propulsion system design. *Journal of Mechanics of Structures and Machines, Automotive Research Center Special Edition Issue*, Haug, E. J. (Ed.), **27**(4), 393–421.

[2] Audet, C., Dennis, J. E. Jr., Moore, D. W., Booker, A. and Frank, P. D. (2000). A surrogate-model-based method for constrained optimization. *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, Paper No. AIAA-2000-4891.

[3] Dixon, L. C. W. and Szegö, G. P. (1978). The global optimisation problem: An introduction *Towards Global Optimisation 2.* North-Holland Publishing Company.

[4] Fellini, R., Papalambros, P. and Weber, T. (2000). Application of a product platform design process to automotive powertrains. *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, Paper No. AIAA-2000-4849.

[5] Fellini, R., Michelena, N., Papalambros, P. and Sasena, M. (1999). Optimal design of automotive hybrid powertrain systems. *IEEE EcoDesign '99 Conference*, Tokyo, Japan.

[6] Gomez, S. and Levy, A. (1982). The tunneling method for solving the constrained global optimization problem with several non-connected feasible regions. In: Dold, A. and Eckmann, B. (Eds.), *Lecture Notes in Mathematics* 909, Springer-Verlag, pp. 34–47.

[7] Goovaerts, P. (1997). *Geostatistics for Natural Evaluation.* Oxford University Press, New York.

[8] Jones, D. R., Perttunen, C. D. and Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Application*, **79**, 157–181.

[9] Jones, D. R., Schonlau, M. and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**(4), 455–492.

[10] National Renewable Energy Laboratory, Department of Energy. [Web Site] URL: http://www.ctts.nrel.gov/analysis/ [Retrieved on 2001, July 4].

[11] Sasena, M. J., Papalambros, P. Y. and Goovaerts, P. (2001). The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. *The Fourth World Congress of Structural and Multidisciplinary Optimization*, Dalian, China.

[12] Sasena, M. J., Papalambros, P. Y. and Goovaerts, P. (2000). Metamodeling sampling criteria in a global optimization framework. *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, Paper No. AIAA-2000–4921.

[13] Schonlau, M, (1997). Computer experiments and global optimization. *Doctoral Dissertation*, Department of Statistics, University of Waterloo.

[14] Watson, A. G. and Barnes, R. J. (1995). Infill sampling criteria to locate extremes. *Mathematical Geology*, **27**(5), 589–608.