

Error Estimation Approaches for Progressive Response Surfaces*

V. J. Romero[†], R. Slepoy, L. P. Swiler, A. A. Giunta
Sandia National Laboratories[‡], Albuquerque, NM

T. Krishnamurthy
NASA Langley Research Center, Hampton, VA

Abstract

Response surface functions are often used as simple and inexpensive replacements for computationally expensive computer models that simulate the behavior of a complex system over some parameter space. "Progressive" response surfaces are built up incrementally as global information is added from new sample points added to the previous points in the parameter space. As the response surfaces are globally upgraded, indications of the convergence of the response surface approximation to the exact (fitted) function can be inferred. Sampling points can be incrementally added in a structured or unstructured fashion. Whatever the approach, it is usually desirable to sample the entire parameter space uniformly (at least in early stages of sampling). At later stages of sampling, depending on the nature of the quantity being resolved it may be desirable to continue sampling uniformly (progressive response surfaces), or to switch to a focusing/economizing strategy of preferentially sampling certain regions of the parameter space based on information gained in previous stages of sampling ("adaptive" response surfaces). Here we consider progressive response surfaces where a balanced representation of global response over the parameter space is desired. We use Kriging and Moving-Least-Squares methods to fit Halton quasi-Monte-Carlo data samples and interpolate over the parameter space. On 2-D test problems we use the response surfaces to compute various response measures and assess the accuracy/applicability of heuristic error estimates based on convergence behavior of the computed response quantities. Where applicable we apply Richardson Extrapolation for estimates of error and asymptotic convergence, and assess the accuracy of these estimates. We seek to develop a robust methodology for constructing progressive response surface approximations with reliable error estimates.

Introduction and Background

Large-scale optimization and uncertainty analyses are often made feasible through the use of response surfaces as surrogates for computational models that may not be directly employable because of prohibitive expense and/or noise properties and/or coupling difficulties in multidisciplinary analysis. Examples of response surface usage to facilitate large-scale optimization and uncertainty/sensitivity analyses are cited in Giunta *et al.* (1994), Roux *et al.* (1996), Unal *et al.*, Venter *et al.* (1996), Romero (1998), and Simpson *et al.* (2004).

Some issues that arise when using response surface approximations (RSA) are accuracy and the number and placement of the data samples on which the RSA is built. With a sufficiently flexible global fitting/interpolating function over the parameter space, response surface accuracy ideally increases as the number of data points increases (if the points are

*This paper is a work of the United States Government and is not subject to copyright protection in the U.S.

[†]Corresponding author, contact: vjromer@sandia.gov, 505-844-5890

[‡]Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

appropriately placed throughout the parameter space), until the essential character of the function is effectively mapped out. Thereafter, it is not cost effective to continue adding samples. Since a single high-fidelity physics simulation (*i.e.*, one data sample) can take many hours to compute, it is desirable to minimize the number of simulations that are needed to construct an accurate response surface.

It can be very difficult to determine when a particular sampling design and interpolation scheme sufficiently resolve a function, yet this must be done if the response surface is to be used as an effective economical replacement for the actual function. Monitoring convergence heuristics of progressive response surface approximations can help in this regard, and to get error estimates once the sampling is stopped. The latter is one of the principle investigative points of this paper.

For our purposes here it is assumed that: 1) the computer model is relatively expensive to evaluate; 2) the parameter space is a unit hypercube or can be accurately and inexpensively mapped into one; 3) the sampled or “target” function is a continuous, deterministic function over the parameter space; 4) reasonably general, arbitrary target functions are to be fitted; and 5) approximate response values are desired over the entire parameter space being considered –*i.e.*, for global and local optimization or mapping inputs to outputs in uncertainty propagation.

Given these specifications, Romero et al. (2004a) examined several data fitting/interpolation methods (kriging, global regression polynomials, and finite-element local polynomial interpolation) for constructing progressive response surfaces built on Progressive Lattice Sampling (PLS) incremental sampling designs. PLS is a paradigm for structured uniform sampling of a hypercube parameter space by placing and incrementally adding sets of samples such that all samples are efficiently leveraged as the design progresses from one level to the next (Romero and Bangston, 1998). Familiar structured designs such as Box-Behnken and Central Composite designs exist or are added at various stages or levels of the design.

PLS endeavors to preserve uniformity of sampling coverage over the parameter space in the various stages of the incremental experimental design. Uniform coverage over the parameter space is desirable for general response surface construction because this reduces the redundancy or marginalization of new information from added samples. This is a basic concept of upgradable quadrature methods (Patterson-1968, and Genz & Malik - 1983).

PLS builds knowledge by reducing global knowledge deficit over the parameter space. It does not attempt to build specific or targeted knowledge by building on previous information in the manner of “adaptive” sampling, which efficiently maximizes knowledge over particular regions of the parameter space. Thus, PLS designs select sample locations strictly on geometric principles such that new samples are intended to be “maximally far” from each other and from all other existing samples at each level of sampling. The locations of previous samples are respected because it is desired to fully leverage them (with minimal marginalization of their information value) as new samples are added. This also means that the new added samples have maximal information value as well.

The advantage of the structured PLS approach is that it is thought to be close to an optimal incremental sampling method in that, as samples are added to attain each new Lattice level, the spacing of samples throughout the parameter space remains uniform. Thus, glo-

bal uniformity of coverage is maintained at each level as the sampling levels progress. This would appear to be the most efficient way (following the precedent of upgradable quadrature methods) to progressively build up a response surface.

Disregarding the constraint of previous or future sample locations, if complete freedom were allowed where the samples could be placed, then these could be arranged over the parameter space to give better results than PLS arrangement provides at any given sampling level. For example, Romero et al. (2003) found that this sometimes occurs with Latin Hypercube Sampling (LHS) and Centroidal Voronoi Tessellation (CVT). Latinized CVT and the Hammersley “Quasi- Monte Carlo” (QMC) low-discrepancy sequence methods also yield very effective point sets (see Romero et al. 2005).

However, the four methods just cited are **non-incremental sampling methods**; augmenting the number of samples implies a completely different sampling of the parameter space with all new point-locations. To go from N to $N+M$ samples in the space would therefore require $N+M$ new evaluations of the exact function (instead of just M new samples). This is more expensive than an **incremental sampling method** like PLS, which at each new stage costs only the increment of M new samples.

A substantial *disadvantage* of PLS, however, is that its **structured** experimental-design nature allows only a quantized increment M of samples to be added to an existing PLS level (point set) to graduate to a new level. Hence, there is a constraint on the number of samples that can be added at a time to maintain the uniform filling of the hypercube. Unfortunately, this quantized incremental cost M accelerates quickly as the PLS level and dimension of the parameter space increase.

Other incremental sampling methods exist that are **unstructured** and do not have the prohibitive cost-scaling of PLS. Some popular unstructured incremental sampling methods are the Halton QMC low-discrepancy sequence method that we examine later in this paper, and Simple-Random Sampling (SRS) Monte Carlo. These allow M as small as 1 without creating a prejudiced imbalance in the global coverage of the parameter space. Hence, regardless of the dimension of the space and the stage of sampling, we could incrementally add, say, 20% more samples at a time and the point sampling would remain as uniform as the method is capable of producing. The question is: “How uniform is the point spacing of these methods?”

The point placement of Halton sampling appears characteristically more uniform over the parameter space than that of SRS, but not as uniform as PLS or the non-incremental sampling methods (see Romero et al. 2004a,b). Among the incremental sampling methods, Romero et al. (2004b) confirm that better sampling uniformity over the parameter space strongly correlates with better response surface accuracy (PLS most accurate, then Halton, then SRS). Because Halton does not suffer from the cost-scaling problems that PLS does, and it generally produces better point uniformity than SRS, we further examine Halton sampling in this paper as a basis for progressive response surfaces.

Given a set of sampling points over a parameter space, the quality of the response surface approximation also depends on the particular method used to fit and interpolate the data. We now turn to consideration of data fitting methods to interpolate and extrapolate the sample data.

Finite-element (FE) type interpolation of point data is very robust and accurate compared to other methods (Romero et al. 1998, 2004a,b; Krishnamurthy et al. 2002). However, FE interpolation is presently limited to structured point placement and low dimensional parameter spaces. For general unstructured point placement and arbitrary numbers of samples in arbitrary dimensions, it is not immediately obvious how anything but globally continuous C^0 (only piecewise smooth) interpolation functions from linear simplex elements could be built. Any higher-order convergence potential in the interpolation function would therefore be lost. Also, procedures for extrapolation to the hypercube boundaries are not immediately obvious, as spatial extrapolation is not normally encountered in the Finite Element Method (Strang & Fix, 1973).

Accordingly, four available data fitting and interpolation/extrapolation methods that are applicable to structured or unstructured progressive sampling schemes have been evaluated by the authors. These are global polynomial regression and kriging (Romero et al. 2004a, Krishnamurthy et al. 2002); Moving Least Squares (MLS) methods (Krishnamurthy et al. 2002, Romero et al. 2003, 2005); and Radial Basis Function (RBF) methods (Krishnamurthy 2003). The MLS, RBF, and kriging methods are capable of yielding much better results than global polynomial regression. However, the former methods are more complex and tricky to use. We see some evidence of this later in the paper. We used only kriging and MLS methods here because we are more familiar with these than with RBF methods, and global polynomial regression is not considered to have the local fitting conformability we desire here.

Finally, we consider the issue of error estimation in response surface approximations. Upon upgrading a response surface by addition of samples, the resulting change in values calculated from the response surface (such as computed optima or volume integrals) is a heuristic indicator of the magnitude of *relevant* error in the response surface approximation. When the incremental change goes to zero (for a sufficiently large number of added samples), this tentatively indicates that the *relevant* response-surface error has become negligible. We elaborate on this basis for error indication in the next section of this paper, and in the subsequent section examine the accuracy of derived error estimates. We also examine whether Richardson Extrapolation, which is a familiar method for assessing the convergence of solution results from successively refined discrete approximations of partial differential equations (Roach 1998), can be successfully used to assess the convergence of results from successively refined response-surface approximations. In doing so, we summarize the RE technique and provide an analogy for its applicability to progressive response surface applications. As we explain below, the Richardson procedure sets constraints on the numbers of samples that can be added to the RSAs at each stage of progression.

2-D Test Problems and Methodology for Assessing Performance of Progressive Response Surfaces and Error Indicators

2-D Test Function

Figure 1 plots the 2-D model function used to study the effect of sample point addition, sample placement, and interpolation method on response surface accuracy. The multimodal function is defined as:

$$Y(p1,p2)=\left[0.8r+0.35\sin\left(2.4\pi\frac{r}{\sqrt{2}}\right)\right][1.5\sin(1.3\theta)] \quad (1)$$

where $r = \sqrt{(p1)^2 + (p2)^2}$ and $\theta = \arctan\left(\frac{p2}{p1}\right)$ on the domain $0 \leq p1 \leq 1, 0 \leq p2 \leq 1$.

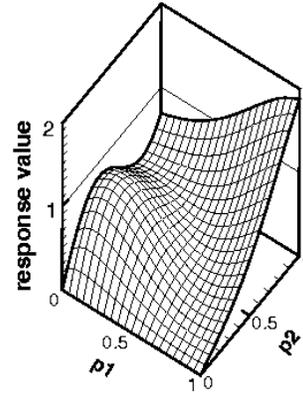


Figure 1. 2-D test function to be approximated (exact “target” function).

Generalized Richardson Extrapolation Methodology

Values (samples) of this target function are obtained at 3, 9, 27, and 81 points to stay within a progression that will support generalized Richardson Extrapolation (Roach 1998) with a grid refinement factor $r = \sqrt{3}$ in each of the two space dimensions $p1$ and $p2$. This factor is squared (would be cubed for three space dimensions, etc.) to determine the growth in the number of subdivisions that the 2-D domain must be discretized into for a uniform refinement of the domain. This 2-D growth rate being three, then starting with 3 subdivisions of the domain (for reasons explained below) the next refinement is to $3 \times 3 = 9$ subdivisions, then $3 \times 9 = 27$, then $3 \times 27 = 81$.** We pick $\sqrt{3}$ as the refinement factor in this work for reasons explained later.

The presumptions here are that at each stage of sampling ($N = 3, 9, 27,$ and 81 points), the N sample points are approximately uniformly spaced throughout the 2-D domain and associated with each one is $1/N$ of the volume of the domain. If the sampling method happens to be Centroidal Voronoi Tessellation then these presumptions hold exactly, and furthermore each sample point is at the volume center of the subdomain (Voronoi “cell” or “region”) associated with the point. These conditions hold less well for less uniform sampling methods like Halton QMC, but still hold approximately. We start with 3 subdivisions here because the associated 3 points are the minimum number in 2-D space that can support a linear response function over the space. This is the lowest-order representation we care to consider with our Moving Least Squares interpolation scheme.

Thus, we have some basis for a discretization-refinement analogy between domain refinement in our response surface progressions and domain refinement in grid solution of partial differential equations (a more familiar context in which we are accustomed to thinking about and applying Richardson Extrapolation). The other element we need to consider is rate of convergence of computed results (like statistics of response behavior computed from progressive RSAs) to the asymptotic exact values that would be obtained in the limit

**This is analogous to a refinement factor of 2 in each space dimension, associated with a 4-times reduction in cell volume in 2-D space when halving of the point-to-point mesh spacing in each coordinate direction, e.g., in going from 2×2 cells to 4×4 cells when each cell is subdivided into 4. The growth factor is then from $2 \times 2 = 4$ cells to $4 \times 4 = 16$ cells, which is a growth from 4 to 16 cells or a growth factor of 4 in two dimensions or $4^{1/2} = 2$ in each dimension.

of infinite refinement. Here, our analogy with grid solution of partial differential equations is even more tenuous. First, only with robust interpolatory methods like low-polynomial-order finite-element interpolation can we be assured that the exact solution is converged to as the number of sample points increases.^{††} The kriging and MLS methods we apply here are not guaranteed to behave well when sample points are added (which we will see in some examples later). Nevertheless, if we can get these methods to work well, can the Richardson Extrapolation idea work out? This depends on whether the quantity being computed with the RSA converges at a constant rate as the domain discretization is refined at a constant refinement rate r per domain dimension. We have not performed the analysis to determine whether theoretically this is the case for MLS or kriging, but we empirically test whether this seems to be the case.

We can empirically solve for an apparent rate of convergence if we have results from three successive discretizations of the domain and the results are monotonically converging at a slowing pace emblematic of the “asymptotic region” of convergence of the discretized solution. We will focus more on these conditions in some examples later. For now assuming that the conditions are satisfied, the following is the standard formula (Roach 1998) for calculating the empirical rate of convergence p :

$$p = \frac{\ln\left(\frac{T_c - T_m}{T_m - T_f}\right)}{\ln(r)} \quad (2)$$

where in this case the per-dimension refinement factor is $r = \sqrt{3}$ and T_c , T_m , and T_f refer respectively to results calculated from RSA built on the coarse, medium, and fine samplings of the exact function. We emphasize here that RE is applied to quantities calculated from the coarse, medium, and fine response surfaces, to improve these estimates or to get error estimates. RE is not applied to improve the response surfaces themselves.

The observed p order can now be used to extrapolate from the two most accurate results (from the medium and fine samplings) to a converged result. The general p^{th} order extrapolation is given by

$$T_{exact} \approx T_f - \frac{T_m - T_f}{r^p - 1}. \quad (3)$$

Our choice of $\sqrt{3}$ as the per-dimension refinement factor in this work arises from several considerations. First, we want to be able to test the accuracy of Richardson Extrapolation and the consistency of the calculated rate of convergence p across two “spans” of response data. That is, we hope to apply the equations over the 3-9-27 point span and over the 9-27-81 point span to see if the calculated convergence rates remain constant, and whether extrapolated results from these two spans of data concide closely. To get the necessary four refinement levels, the augmentation factor of $\sqrt{3}$ per space dimension or $\sqrt{3}^2 = 3$ for the two dimensions yields a reasonable number of top-end samples (81) in a progression that lasts 4 levels. This somewhat arbitrary choice simply allowed us to take a reasonable initial look at the possibilities here.

^{††}A mathematical analysis of finite element interpolation (Strang and Fix, 1973) shows that for a continuous and infinitely differentiable function over the parameter space, the domain integral of the pointwise absolute error goes to zero as the spacing between samples goes to zero.

Error Estimation by “Simple Delta” Method in Progressive RSA Calculations

A very simple error estimation approach applicable to progressive RSA results is tied to the change in computed response values as samples are added. This “**simple delta**” method of error estimation is patterned after the control of convergence error in many fields such as local optimization, root-finding, and the iterative solution of sets of nonlinear simultaneous equations (often associated with the solution of partial differential equations, *e.g.*, P/Thermal Theory Manual, 1991).

In these applications it is necessary to determine when to stop iterating. Upon completing an iteration (or the addition of a sufficiently large number of samples to the RSA), if the incremental change in the quantity being computed is exactly zero, this generally indicates that the iteration has converged (although there are imaginable situations where this would not be the case). However, it is generally not cost effective to require machine-zero change in the computed quantity before terminating the iteration. Instead, an acceptably larger error tolerance is specified that terminates the iteration when the incremental change in the computed quantity becomes smaller than the tolerance. The tolerance is set according to how much error from the true result the user is willing to accept, in the context of the tradeoff that more accuracy implies more computing cost. The hope here is that when the incremental change becomes less than the prescribed tolerance, the answer is within that tolerance of the exact result.

This simple method of controlling and estimating convergence error appears to work reasonably well in practice, but there are certainly many circumstances in practice where it will fail. In particular, if the tolerance condition is met in a slowly-converging regime of the iteration process, then termination can be premature and the exact result may not be within the specified error tolerance. Techniques are therefore sometimes employed to account for the actual rate of convergence (fast or slow, whatever it may be), but these are usually less-sophisticated variations of the RE procedure described above. We therefore assume that our empirical results with the RE method (presented later in this paper) represent the best that can be practically attained when convergence-rate information is factored into the error estimates.

We note an important difference between the convergence-error control procedure and our *simple-delta* error estimation method. A large body of empirical evidence reveals that in the convergence-error control process the true error is frequently successfully bounded by the specified error tolerance value. However, this tolerance can be much larger than the actual iterative delta that first falls below this tolerance and stops the iteration. The existing empirical evidence that the tolerance value has been a fairly reliable error bound does not mean that the smaller actual iterative delta (our simple-delta error estimate) will likewise routinely bound the actual error. Therefore, we must rely on our empirical study here to give us first indications of whether the simple-delta method might work with reasonable reliability.

Halton Point Sets

With the general methodology explained, we proceed to the specific conditions of our numerical experiments. For the reasons cited earlier, we use Halton QMC sampling (see *e.g.* Owen 2003) to generate the 3, 9, 27, and 81 point progressions. We use two different sets of prime-number bases (2-3 and 5-7) to create two different sets of Halton point progressions to determine any effect of prime-number base choice on the results. Figure 2

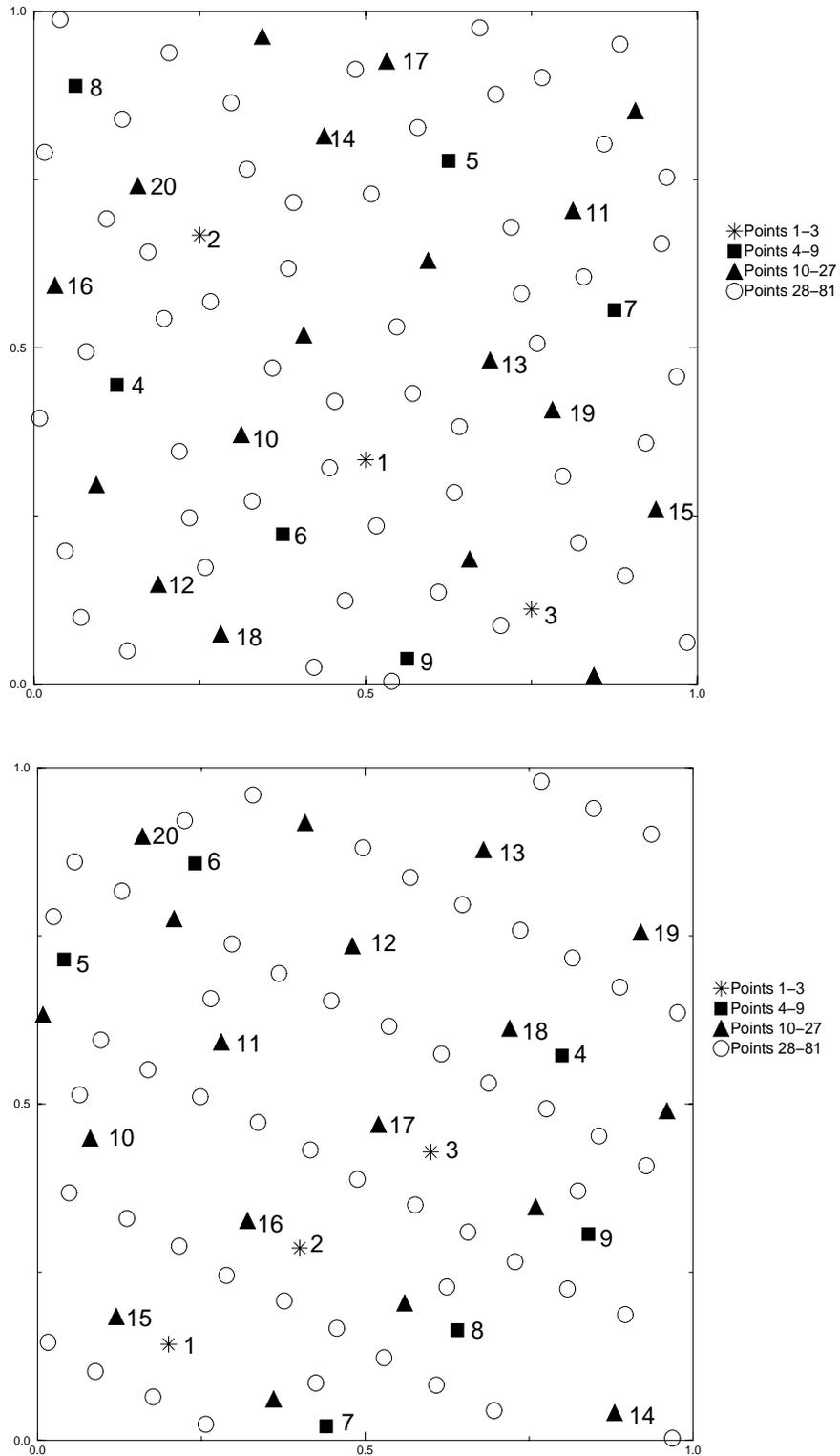


Figure 2. Halton Quasi-Monte-Carlo point sets (3, 9, 27, and 81 point sets from two different prime-number bases (2-3 [top chart] and 5-7 [bottom chart])). First 20 consecutive points in each progression are labeled to observe the uniformity (or lack thereof) in point spacing as points are added one at a time.

shows the specific Halton point sets used. The first 20 consecutive points in each progression are labeled so that the reader can see the uniformity (or lack thereof) in point spacing as sampling points are added one at a time.

Kriging and Moving Least Squares Data Fitting/Interpolation Approaches

The point data is fitted and interpolated in the following with Moving Least Squares (MLS) and kriging methods. The particular implementation of MLS we use is described in Krishnamurthy et al. (2002). Linear basis functions are used for the 3-point Halton 2-3 and 5-7 sets. For the 9-point and larger sets a quadratic polynomial basis function is used which requires at least $(D+1)(D+2)/2$ sample points (6 for $D=2$ dimensions) within a given interpolation point's local radius of influence. An optimal local radius of influence is calculated and used for each different point set so that this element of fitting error is minimized in this study. Optimal influence radii are determined by iterating the radius value to minimize the sum of the squared differences between the particular response surface and the exact function at a 21×21 grid of uniformly space points over the 2-D domain. A quintic weighting function was used to give C^2 smoothness to the MLS interpolation function over the 2-D domain.

The kriging method used in this study also produces a continuous C^2 interpolating function over the entire parameter space. The fundamentals of the method are described in (Giunta et al. 1998). The method is principally affected by the correlation length parameter Θ of the exponential spatial correlation model in the method. The correlation length parameter can be specified or computed internally using a maximum-likelihood-estimation optimization procedure. Both approaches were tried as described in the next section.

Performance of Progressive Response Surface Methods and Error Indicators

Figure 3 shows initial kriging response surfaces based on letting the code internally calculate the spatial correlation length parameter Θ . The method did not return results (due to ill conditioning) at 81 points for either the base 2-3 or 5-7 Halton sets. Furthermore, the base 2-3 3-point and base 5-7 3-point and 9-point response surface approximations seemed suspect based on visual appearance.

We then switched to the input option for specifying the value of Θ and proceeded to iterate the value until the ill-conditioning problem was eliminated for the 81-point sets and the results were fairly stable to local perturbations in the Θ value. Figure 4 shows the resulting RSAs, along with the results from using the MLS approach. Generally, the progressions of the externally optimized MLS and kriging methods are orderly toward the best representations at 81 points. Beyond the noticeable differences at 3 points, the differences between kriging and MLS are relatively small at 9, 27, and 81 points. The effect of prime-number base is significant at lower numbers of samples (3 and 9 points) but not for the larger sample sets (27 and 81 points) when the parameter space becomes well populated.

The MLS results at 3 points are noticeably bad for both 2-3 and 5-7 sets. This is because a linear basis function is the only allowable MLS surface for 3 points in a 2-D space and the first three points that define the associated plane are almost collinear for base 2-3 and are collinear to within machine precision for base 5-7 (see Figure 2). Thus, the 5-7 case is ill-conditioned so the MLS method yielded noise. The base 2-3 results are a steep plane (though shown in Figure 4 with flattened dog-ear extremes because of plotter limitations

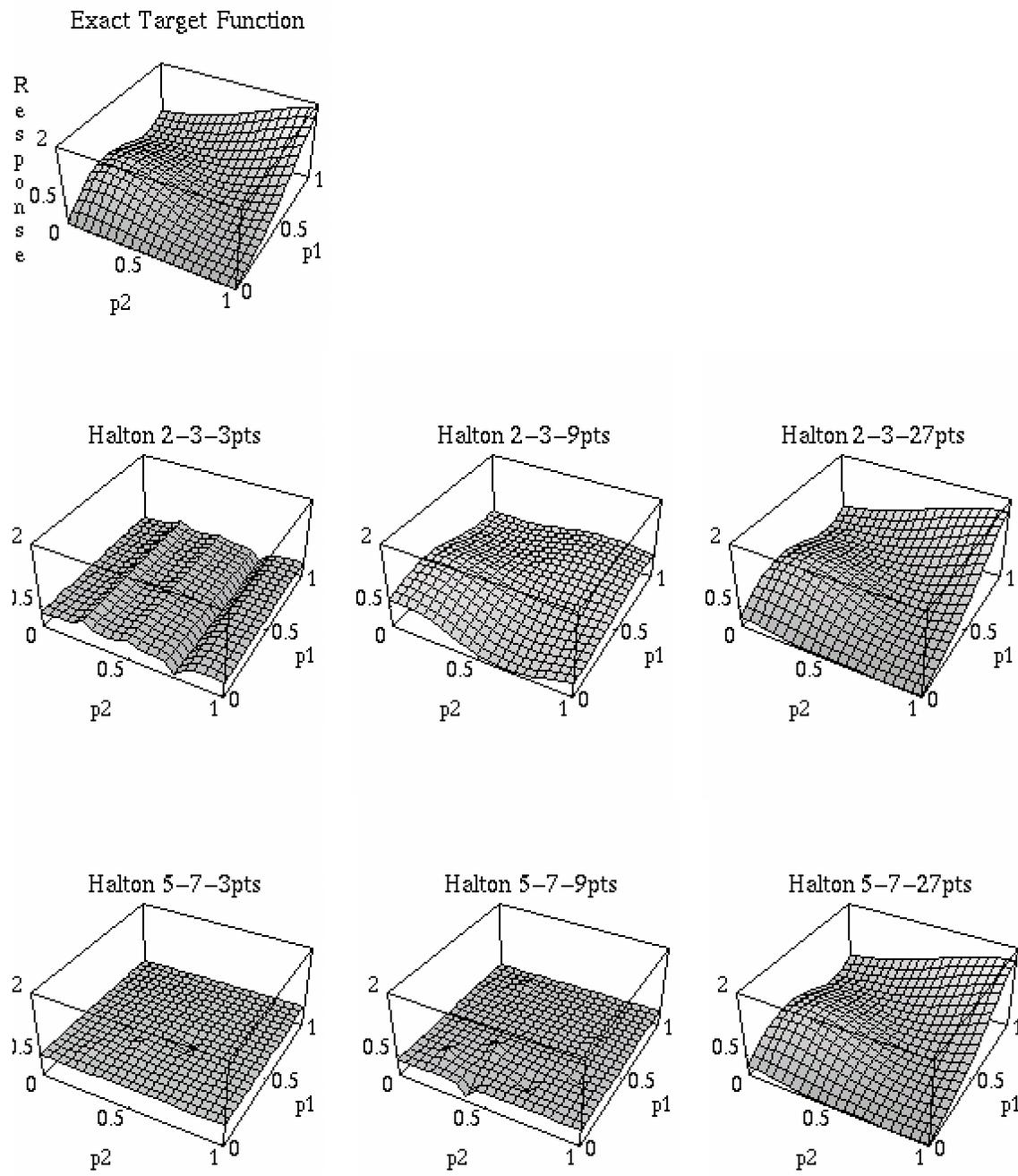


Figure 3. Initial kriging response surfaces for Halton base 2-3 and 5-7 sample sets (maximum-likelihood optimization problem internally solved for spatial correlation-length parameter Θ).

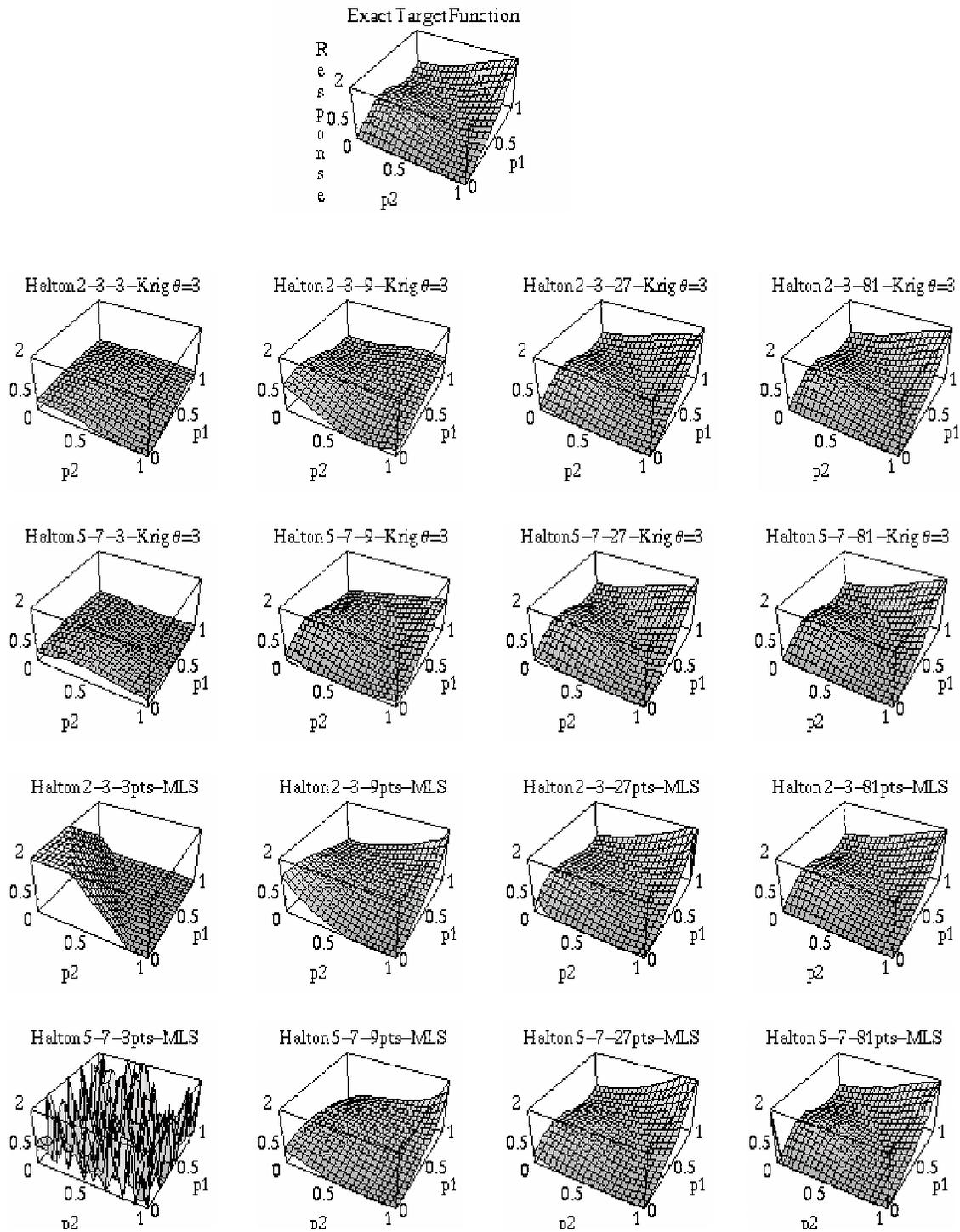


Figure 4. Convergence of progressive response surfaces to target function as samples are added to the parameter space. (Kriging and MLS spatial correlation length parameters optimized externally by comparison to exact target function.)

when the vertical scale was kept the same as the other plots on the page). We could have legitimately eliminated the MLS problems at 3 points by simply picking 3 best points out of the sets of 9 (assuming that in a real application we would plan to progress to at least a second level of sampling and would therefore know *a priori* nine locations for sample points and be able to pick the best three of these locations to sample first). However, to be fair we would have had to use the same new base 2-3 and 5-7 3-point sets for the kriging progressions and had already calculated many results with the original data (presented next). Since the unfortunate MLS results with the original 3-point sets don't adversely impact the error estimation findings below, the resulting response surfaces are retained.

Spatially Averaged Pointwise Error of Response Surface Approximations

A simple measure of global RSA accuracy over the parameter space can be calculated as:

$$\text{approximate spatial average pointwise error} = \frac{\sum_{i=1}^{441} |\text{exact}_i - \text{predicted}_i|}{441} \quad (4)$$

where exact and predicted values in the summation come from respective evaluation of the exact function and the particular RSA at 441 uniformly spaced points on a 21x21 square grid overlaid on the domain. This measure is an expedient approximation to the global average integrated absolute error over the domain, which would require a much more involved calculation. The value of our metric can depend strongly on the grid used if it under-resolves the functions. As Figure 4 suggests, a 21x21 plotting grid appears to be sufficiently dense to achieve adequate representation of the target and approximation functions.

Figure 5 presents results from the various interpolation methods and Halton point progressions. The MLS results show orderly reduction of global fitting error as the number of Halton samples increases. (Error at 3 points is off the scale of the plots because of the unfortunate oversight described earlier where the three points used were effectively collinear. This is viewed as an avoidable interaction with the Halton sampling method.) The kriging results generally also show orderly error reduction except that the 81-sample result in the Halton base 5-7 progression has greater average absolute error than the 27-sample result. We attribute this to possible degraded conditioning of the kriging solution matrix at the larger allotment of sample points. (Ill conditioning prevented solutions at 81 points for both 5-7 and 2-3 sets under the Θ values initially tried, Figure 3.) Nonetheless, the MLS and kriging values are closely comparable at 9, 27, and 81 sample points of each Halton progression (base 2-3 and base 5-7).

Accuracy of Global Statistics of Response Calculated from RSAs

Global RSA accuracy as it impacts uncertainty propagation will be assessed next. In a probabilistic uncertainty propagation problem, the system response function maps uncertain (random variable) inputs of the system into an output uncertainty distribution. Concretely, let the values of two uncertain inputs $p1$ and $p2$ come from independent normal distributions having means 0.5 and standard deviations $\sigma=0.5/3$. The corresponding joint probability density function (JPDF) of these input random variables is represented in Figure 6 after truncation of the JPDF beyond the $p1$ - $p2$ unit-square domain. The JPDF is then renormalized to integrate to unity over this space. The relative height of the JPDF defines the relative propensity for attaining specific combinations of input values marked by any

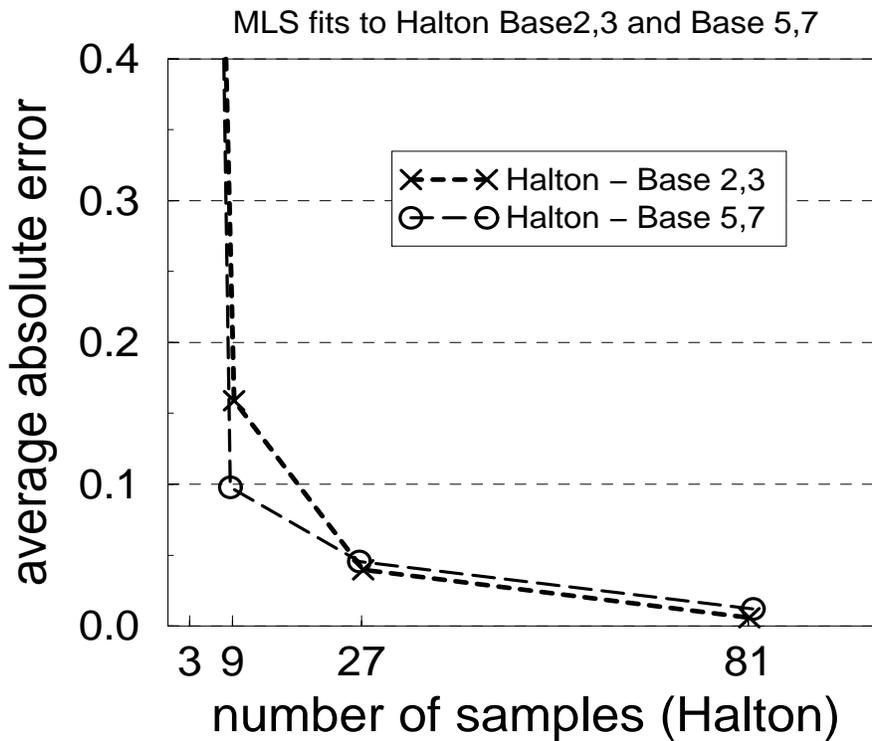
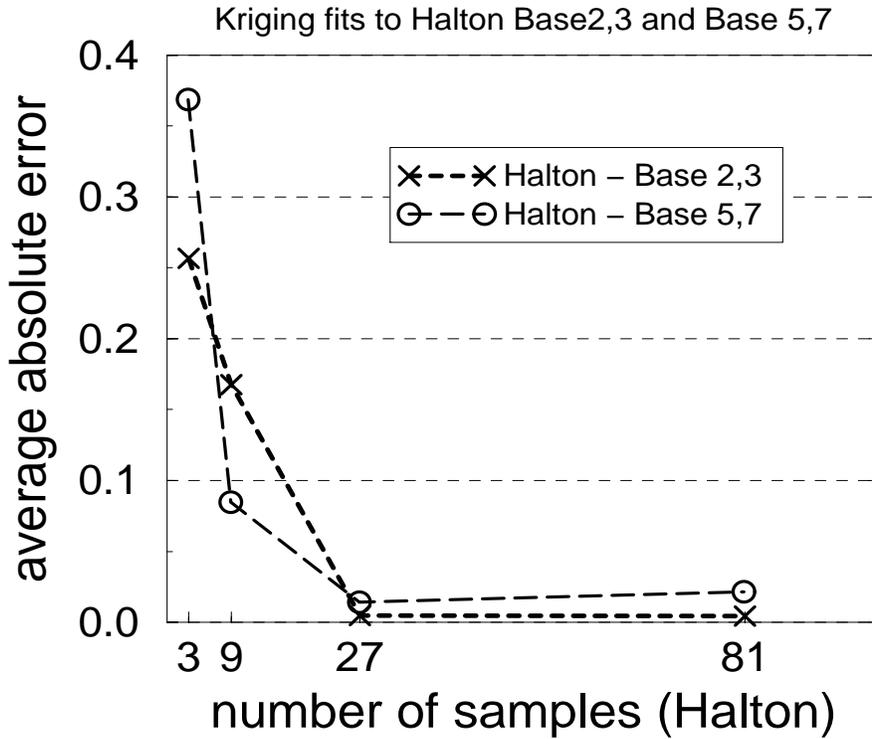


Figure 5. Behavior of spatially averaged pointwise absolute error as sample points are added in kriging and MLS response surface formulations.

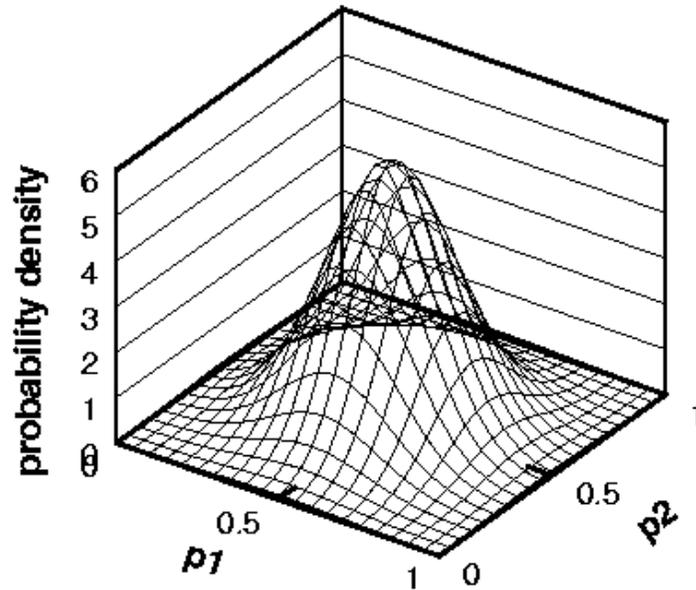


Figure 6. Joint Probability Density Function describing the random variables in the problem: normally distributed parameters $p1$ and $p2$ with means 0.5, std. deviations $\sigma=0.167$, and truncation of the unit square parameter space at 3σ above and below the mean values.

particular point (location) in the parameter space. This likelihood function for attaining given input combinations maps through the applicable input/output functional relationship of the system (such as our test function shown in Figure 1) into a corresponding likelihood function for attaining values of the system output.

Operationally, the JPDF of the input random variables can be mapped through the system input/output response function into an output response distribution via standard Monte Carlo (MC) sampling techniques. Toward this purpose, one million LHS MC samples or $(p1, p2)$ parameter combinations (points in the parameter space) were generated by the Iman and Shortencarrier (1984) LHS code according to the JPDF in Figure 6. The 2-D target function and the various RSAs in Figure 4 were then evaluated for 10^6 $(p1, p2)$ parameter sets, yielding a population of 10^6 response values for each function.

The impact of the response surface constructions on the accuracy of global statistics calculated from the mapped response values are shown in Figures 7 and 8. Figure 7 shows calculated means of 10^6 response values calculated respectively from response surfaces constructed from 3, 9, 27, and 81 Halton points. Figure 8 shows calculated standard deviations. Also shown for reference are the exact values of the statistics obtained from the exact 2-D function. Much more so than for average absolute error of the response surface (Figure 5), the convergence behavior of computed global statistics of response can be quite erratic, as we have found before with PLS sampling schemes (see Romero et al. 2004a).

Calculated Mean of Response

Richardson Extrapolation (RE) is not applicable in cases of oscillatory convergence like with calculated mean response from kriging with the Halton base 2-3 sampling (Figure 7). On the other hand, the Halton base 5-7 results spanning 9-27-81 points do not oscillate,

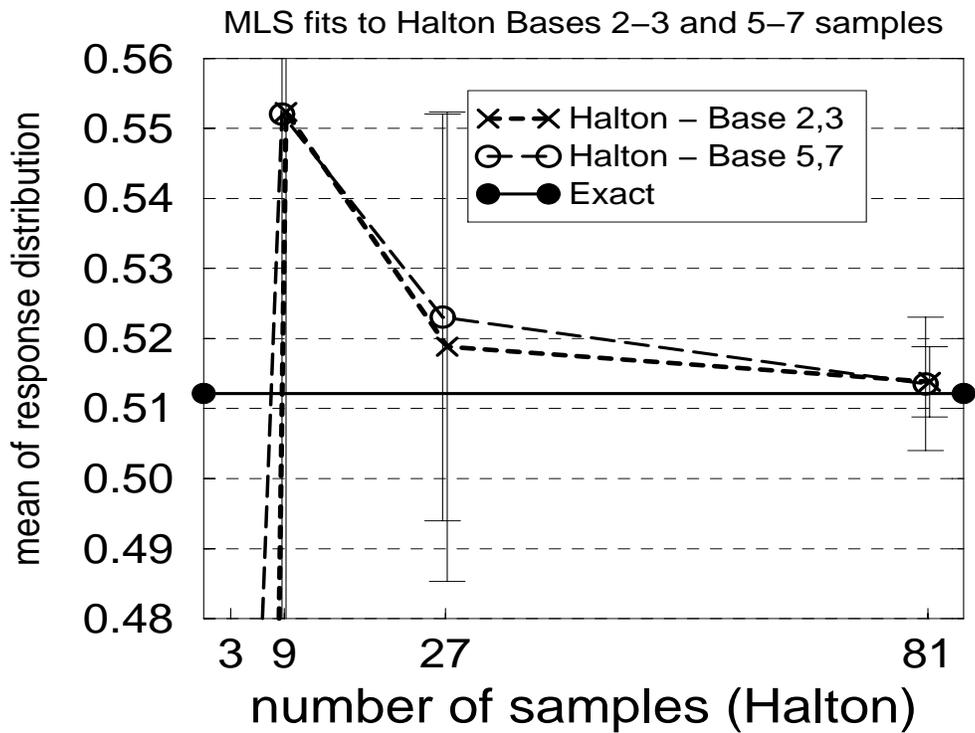
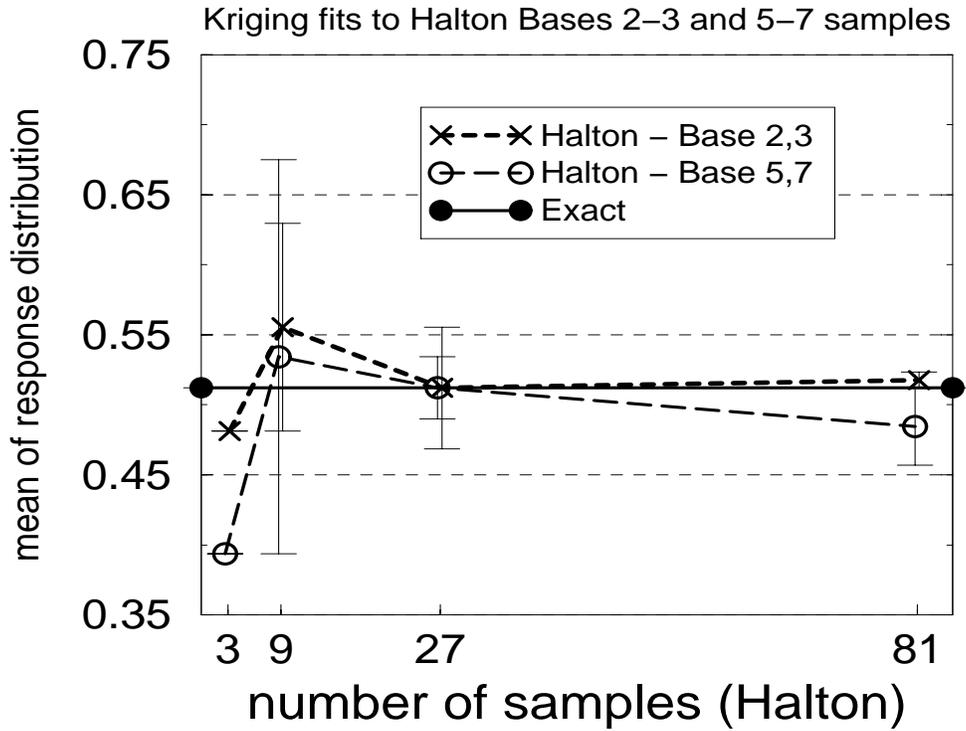


Figure 7. Convergence behavior of calculated mean as sample points are added in kriging and MLS response surface formulations.

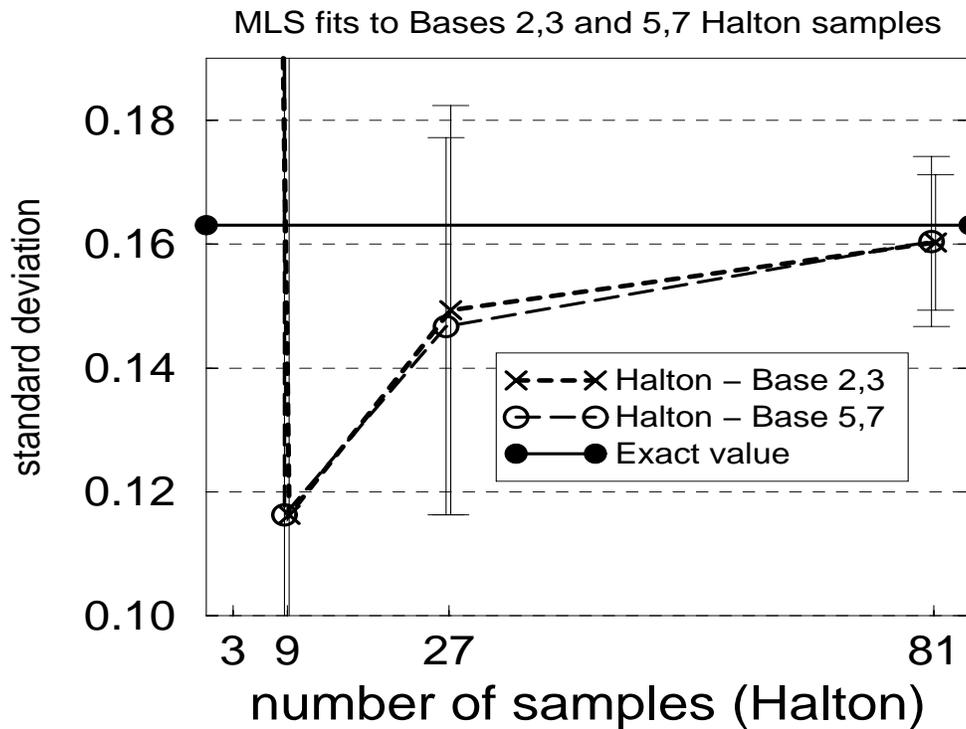
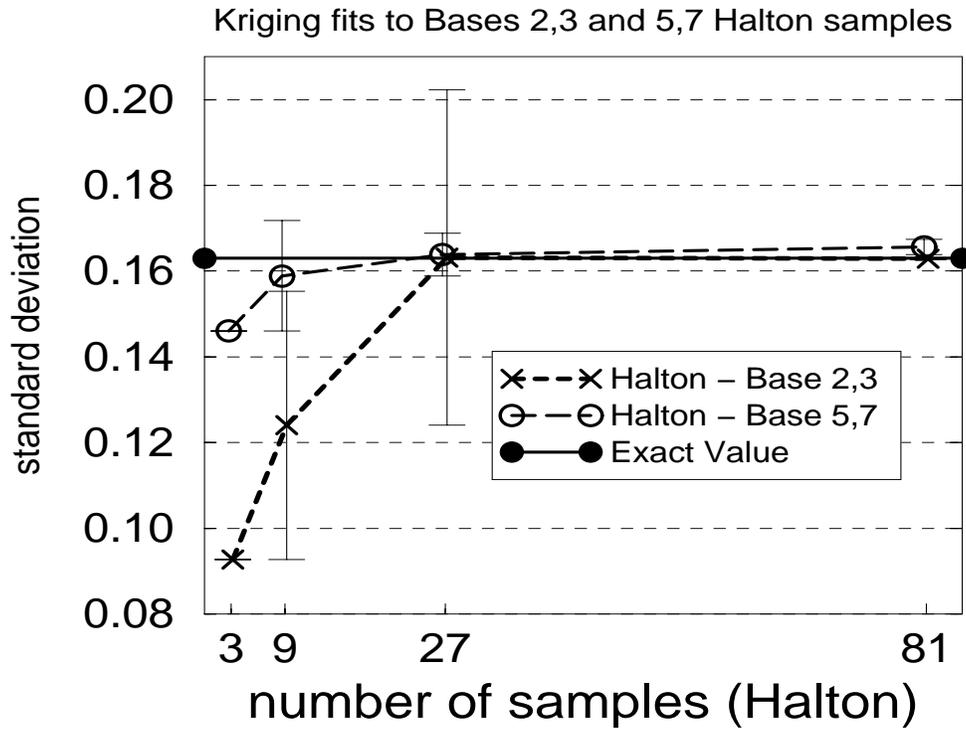


Figure 8. Convergence behavior of calculated standard deviation as sample points are added in kriging and MLS response surface formulations.

but these are also not amenable to RE because the monotonic behavior is obviously not convergent yet because the difference from the 27-point result to the 81-point result is greater than the difference from the 9-point to the 27-point result. Therefore, the behavior is not yet in the asymptotic regime of the response-surface approximation.

Nevertheless, for either behavior above we can apply the more crude and simplistic *simple-delta* method of error estimation by estimating the error bound as the change in results that occurs upon adding the new samples. For example, consider the results in Figure 7. For each curve we can derive error bars at 9, 27, and 81 points by subtracting the preceding coarser result. For instance, we subtract the result at 3 points from the result at 9 points. The absolute magnitude of this difference is taken as an estimate of how much more the exact result may differ from the current value (at 9 points in this example). In view of the often oscillatory nature of convergence here, we form error half-bars of this magnituded above and below the result at 9 points. In this same manner we form the error bars about the results at 27 and 81 points as well.

In all cases in Figure 7 (at 9, 27, and 81 points and for all four curves –kriging with Halton 2-3 and 5-7 sampling progressions and MLS with both Halton progressions), the exact result indeed lies within the error bars so constructed. Thus, the simple-delta approach for calculating error bars worked very well here: the exact mean was contained within all 12 constructed error bars.

We can compare this success with RE error estimates that *are* derivable for the *MLS* results in Figure 7. As explained earlier, we used a different basis function (linear) for the 3-point RSA from that used (quadratic) for the 9, 27, and 81 point RSAs. Hence, we would expect the order of convergence of the Halton-MSL RSA scheme would surely change between the 3-point results and the other results. Therefore, we do not consider the 3-9-27 point span to be amenable to RE even if much better results were obtained at 3 points by advantageous selection of the 3 starting points via the approach explained earlier. However, we might hope that the convergence rate over the 9-27-81 point span is relatively constant (we present some empirical evidence for this later). Applying the RE scheme we determine the Richardson extrapolations of the base 2-3 and base 5-7 progressions. (The extrapolated values are not shown in the plots). In one case (base 2-3 progression), the RE result is closer to the exact value than the most refined calculated result (at 81 points), but for the 5-7 progression the converse is true. In both cases the RE results are within error bars of the exact mean, where the bar half-heights are given by differences from the 81-point values. Hence, the simplistic error estimation approach was again successful here.

We can alternatively take the view that the actual calculated result at the finest resolution is a better estimate than the RE value. Because we have not established a theoretical rate of convergence for our RSA methods, we cannot be assured that the rate should stay constant over our sampling progression and therefore that the RE estimate should be exact. Empirically, we have already seen above that the 81-point value is more accurate than the RE value in one of the two cases. Hence, we may consider the 81-point value to be our best estimate and construct an error bar around it by differencing from the RE value. In this context, the RE machinery is just a relatively complex way to get a simple-delta error estimate on the most-refined (81 point) RSA estimate. We call this the “**RE delta**” method of getting an error estimate. Numerically, the simple-delta error bars about the 2-3 and 5-7 81-point results contain the exact value in both cases, whereas the RE-delta approach only

works in one of the two cases. Hence, here the simpler error estimation approach was successful more frequently.

Calculated Standard Deviation of Response

We next consider the standard-deviation estimates in Figure 8. Addressing the kriging results first, we see that the simply-derived error bars worked in 4 of the 6 possible opportunities. Regarding Richardson Extrapolation, the 3-9-27 point span of the Halton 2-3 progression does not qualify because the monotonic behavior is not asymptotic. Behavior over the 9-27-81 span is oscillatory, so this span does not qualify either. However, both point spans of the 5-7 progression qualify for RE. The calculated rate of convergence for the 3-9-27 span is $p=1.73$, and for the 9-27-81 span is $p=1.88$, so the convergence rate is reasonably constant here, suggesting that RE might work well here. However, it turns out that the 27-point RSA result is more accurate than the asymptotic extrapolation of the 3-9-27 span, and the 81-point RSA result is better than the asymptotic extrapolation of the 9-27-81 span. Here, the RE estimates were worse than the most refined RSA estimate in the progressions from which the asymptotic (RE) estimates arose.

Nonetheless, if we choose to use the RE estimates as our best estimate (in real application settings not knowing the answer, so not knowing which estimate is better), then the exact mean lies outside of the available error-bar estimates (by simple delta) for the RE results. If we conversely choose to use the RE estimates to define error bars around the 27-point and 81-point RSA results (by the RE delta method), then the error bars about the 27-point result would contain the true mean, but the error bars for the 81-point result would not. In comparison, the error bars from simple differencing with respect to preceding results (at 9 and 27 points, respectively) do contain the exact mean as shown in Figure 8. Hence, not only are the RE estimates (and available error bar estimates for them) usually less accurate than the most refined RSA estimates in the progressions they are derived from, but also the RE estimates are not as good for yielding successful error bars (for the most refined RSA estimate underlying the RE estimate) as simple backward deltas are.

Addressing the MLS results next, we see that the simple backward-derived error bars worked in all 6 of the 6 possible opportunities. Regarding Richardson Extrapolation, both 3-9-27 spans are out of contention again because of the problems with the effectively col-linear 3 points in the 2-3 and 5-7 progressions. However, both Halton progressions of the 9-27-81 span qualify for extrapolation. The calculated rates of convergence are $p=2.01$ for the 2-3 progression and $p=1.45$ for the 5-7 progression. Again, the convergence rate is reasonably consistent. Nevertheless, the RSA estimates are about as accurate (2-3 case) or better (5-7 case) than the asymptotic RE estimates. Again here, on balance the RE estimates were worse than the most refined RSA estimate in the progressions from which the asymptotic (RE) estimates arose.

If we choose to use the RE estimates as our best estimate, then the exact standard deviation lies inside the error-bar estimates from simple backward delta. Conversely, if we choose to use the RE estimates to define error bars around the base 2-3 and 5-7 81-point RSA results (by the RE delta method), then the error bars here also contain the true result. In comparison, the error bars from simple differencing with respect to preceding results (at 27 points) also contain the exact value as shown in Figure 8. So here we see that the available error-bar estimates about the RE results do contain the exact standard deviation, and that RE-derived error bars on the 81-point results were as successful as backward differencing.

Discussion and Conclusion

It can be very difficult to determine when a particular sampling design and interpolation scheme sufficiently resolve a function, yet this must be done if the response surface is to be used as an effective economical replacement for the actual function. Monitoring convergence heuristics of progressive response surface approximations can help in this regard, and to get error estimates once the sampling is stopped. The latter is one of the principle investigative points of this paper.

Initial indications based on our limited observations are that very simple delta-based error estimates (error bars), obtained from changes in RSA estimates from addition of sufficiently large increments of samples to the approximation, are fairly successful error indicators. However, success tendencies may be very problem dependent and certainly depend strongly on the size of the sample increments. For example, would simple-delta error estimates work as successfully under a doubling (say) of sample points, such that the progression would go from the lowest reasonable number (3 points in 2-D) to 6, 12, 24 points, etc? If so, then the slower rate of growth in the sampling progression could perhaps reduce the total number of samples required to decide when the exact function has been sampled enough to satisfy the error requirements in the particular computational task. Certainly, much investigation remains to be done to better characterize the general prospects of simple-delta error estimation in general problems and under other sampling incrementation schemes.

Our investigation found that Richardson Extrapolation estimates are often not as accurate as the last computed value in the progression from which the asymptotic RE estimates are derived. Furthermore, the available error bars about the RE estimates by simple differencing were not as successful in containing the exact values as were similarly-produced error bars about the actual RSA estimates. Moreover, the error-bar estimates derived from RE were: 1) not as readily applicable [only calculable for the third estimate in the progression, and not the previous estimates]; 2) not as simple to compute as the simple-delta estimates; and 3) not as frequently successful as error bars derived with the simple-delta approach. We have also found these trends and conclusions to be true in other work (now being written up) concerning estimates of exceedence probabilities for 2-D problems defined in Romero et al. (2004a). Therefore, all our indications to date, though very limited, are that use of RE for estimation improvement and for error estimates is probably not going to be fruitful for progressive RSA methods of the type we are developing.

We also note that we have intervened to manually set the parameters in the kriging and MLS methods to get the results shown here. In doing so, we have used knowledge of the true response function shape to optimize the spatial correlation length parameters input to these methods. We have intervened in order to ensure favorable conditions for the applicability of RE method where possible, in order to see whether the method might work (even if under the artificially favorable conditions). We have found that even under well-controlled favorable conditions, RE was only mildly successful. In contrast, the simple-delta error estimation approach worked more successfully and in a greater variety of circumstances, and does not need the benefit of the control applied here.

Although kriging and MLS have the potential for superior response surface accuracy (versus, e.g., global polynomial regression), at this point we are not confident that they can be used autonomously in general settings. It appears that much more work needs to be

done to make them more robust and autonomous methods that could be used in automated procedures of surrogate-based optimization and uncertainty propagation.

References

- Burkardt, J., M. Gunzburger, H. Nguyen, J. Peterson, and Y. Saka, (2004) "A new methodology for point sampling I: Uniform sampling in hypercubes" in preparation
- Genz, A.C., and Malik, A.A., (1983) "An Imbedded Family of Fully Symmetric Numerical Integration Rules," *SIAM J. Numer. Anal.*, Vol. 20, No. 3, June.
- Giunta, AA., Dudley, J.M., Narducci, R., Grossman, B., Haftka, R.T., Mason, W.H., Watson, L.T., (1994) "Noisy Aerodynamic Response and Smooth Approximations in HSCT Design," paper AIAA-94-4376, 5th AIAA Multidisciplinary Analysis and Optimization Symposium, Panama City, Fla., Sept. 7-9, 1994.
- Giunta, A.A., and Watson, L.T., (1998) "A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models," paper AIAA-1998-4758 presented at the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, MO, Sept. 2-4, 1998.
- Iman, R.L., and Shortencarier, M.J. (1984), "A FORTRAN77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples to Use with Computer Models," Sandia National Laboratories report SAND83-2365 (RG), printed March 1984.
- Krishnamurthy, T., and Romero, V.J., (2002) "Construction of Response Surfaces with Higher Order Continuity and its Application to Reliability," paper AIAA-2002-1466, 43rd Structures, Structural Dynamics, and Materials Conference, April 22-25, 2002, Denver, CO.
- Krishnamurthy, T., (2003) "Response Surface Approximation with Augmented and Compact Support Radial Basis Functions," paper AIAA-2003-7561, 44th Structures, Structural Dynamics, and Materials Conference, April 7-10, 2003, Norfolk, VA.
- Owen, A. (2003), "Quasi-Monte Carlo Sampling," A Tutorial for SIGGRAPH 2003, Stanford University Statistics Dept., notes available at: <http://www-stat.stanford.edu/~owen/reports/siggraph03.pdf>
- Patterson, T.N.L., (1968) "The Optimum Addition of Points to Quadrature Formulae," *Math. Comp.*, 22, pp. 847-856.
- P/Thermal Analysis Code Theory Manual (1991), Section 4.2.4 Convergence Criteria and Error Estimation, PDA Engineering, Costa Mesa, CA.
- Roache, P.J., (1998) *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, New Mexico.
- Romero, V.J. (1998) "Optimization and Non-Deterministic Analysis with Large Simulation Models: Issues and Directions," paper in ASME HTD-Vol. 357-2, 7th ASME/AIAA Joint Thermophysics and Heat Transfer Conference, Albuquerque, NM, June 15-18, 1998.
- Romero, V.J., and S.D. Bankston (1998) "Finite-Element/Progressive-Lattice-Sampling Response Surface Methodology and Application to Benchmark Probability Quantification Problems," Sandia National Laboratories report SAND98-0567, printed March, 1998.
- Romero, V.J., J.V. Burkardt, M.D. Gunzburger, J.S. Peterson, T. Krishnamurthy, (2003) "Initial Application and Evaluation of a Promising New Sampling Method for Response Surface Generation: Centroidal Voronoi Tessellation," paper AIAA-2003-2008, 44th Structures, Structural Dynamics, and Materials Conference, Apr. 7-10, 2003, Norfolk, VA.
- Romero, V.J., L.P. Swiler, A.A. Giunta, (2004a) "Construction of Response Surfaces based on Progressive-Lattice-Sampling Experimental Designs with Application to Uncertainty Propagation," *Structural Safety*, 26, pp. 201-219.

- Romero, V.J., T. Krishnamurthy, L.P. Swiler, (2004b) "Progressive Response Surfaces," 9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, NM, July 26-28.
- Romero, V.J., J.V. Burkardt, M.D. Gunzburger, J.S. Peterson, (2005) "Comparison of Pure and "Latinized" Centroidal Voronoi Tessellation Against Other Statistical Sampling Methods," accepted for special issue of *Reliability Engineering and System Safety* devoted to Sensitivity Analysis of Model Output (SAMO) Conference, Santa Fe, NM, March 8-11. Conference version of paper available at <http://library.lanl.gov/ccw/samo2004/>.
- Roux, W.J., Stander, N., and Haftka, R.T., (1996) "Response Surface Approximations for Structural Optimization," *6th Multi-Disciplinary Analysis and Optimization Symposium*, Bellevue, WA, Sept. 4-6 (paper AIAA 4042-CP).
- Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N. and Yang, R.-J., (2004) "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion," *Structural and Multidisciplinary Optimization*, Vol. 27, No. 5, pp. 302-313.
- Strang, G., and Fix, G.J. (1973), *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, NJ.
- Venter, G., Haftka, R.T., and Starnes, J.H. Jr. (1996), "Construction of Response Surfaces for Design Optimization Applications," *6th Multi-Disciplinary Analysis and Optimization Symposium*, Bellevue, WA, Sept. 4-6 (AIAA paper 96-4040-CP).