Shifting Data Collection from a Fixed to an Adaptive Sampling Paradigm

Erik L. Axdahl
NASA Langley Research Center
Hampton, VA
Erik.L.Axdahl@NASA.gov

ABSTRACT

For domains where data are difficult to obtain due to human or resource limitations, an emphasis is needed to efficiently explore the dimensions of information spaces to acquire any given response of interest. Many disciplines are still making the transition from brute force, dense, full factorial exploration of their information spaces to a more efficient design of experiments approach; the latter being in use successfully for many decades in agricultural and automotive applications. Although this transition is still incomplete, groundwork must be laid for incorporating the next generation of algorithms to adaptively explore the information space in response to data collected, as well as any resulting empirical models (i.e., metamodels). The methodology in the present work was to compare metamodel quality using a fixed sampling technique compared to an adaptive sampling technique based on metamodel variance. In order to quantify metamodeling errors, a delta method was used to provide quantitative model variance estimates. The present methodology was applied to a design space with an air-breathing engine performance response. It was shown that competitive metamodel quality with lower associated error could be achieved for an adaptive sampling technique for the same level of effort as a fixed, a priori sampling technique.

ABOUT THE AUTHOR

Dr. Erik L. Axdahl is a Research Aerospace Engineer in NASA Langley Research Center's Hypersonic Airbreathing Propulsion Branch. Dr. Axdahl specializes in the design and analysis of high speed air-breathing propulsion systems with an emphasis on modeling and simulation spanning multiple levels of fidelity. He is a Senior Member of the American Institute of Aeronautics and Astronautics (AIAA), a member of the AIAA High Speed Air Breathing Propulsion Technical Committee, a technical area organizer for AIAA Propulsion and Energy, and a graduate of the University of Minnesota – Twin Cities and the Georgia Institute of Technology.

Shifting Data Collection from a Fixed to an Adaptive Sampling Paradigm

Erik L. Axdahl
NASA Langley Research Center
Hampton, VA
Erik.L.Axdahl@NASA.gov

INTRODUCTION

The application of validated, physics-based modeling and simulation of complex systems is a way to improve the fidelity and credibility of engineering processes both during design and concurrent with experimental and in-situ testing. For aerospace systems, model-based systems engineering frameworks (e.g., Robinson and Martin 2008) ease the process of managing an ensemble of discipline-specific (e.g., structures, aerodynamics, propulsion) design and analysis codes, facilitating data communication between them, and post-processing results to arrive at key metrics of interest (e.g., lift, drag, thrust). However, a key bottleneck remains that iteratively running each code can take a significant amount of time, especially combined with the desire to bring high fidelity tools earlier in the design process, even to the conceptual level.

The use of metamodels or response surface methods (Myers and Montgomery 2002) has been a valuable way to incorporate higher fidelity simulation results earlier into the design process. Such methodologies are essentially a sampling scheme, such as design of experiments (DoE), combined with empirical modeling and model adequacy assessments. For air-breathing hypersonic aerospace systems, DoE has played a key role in the success of programs of record (McClinton et al. 2002; Bynum and Baurle 2011). However, research and programmatic pressure exists to increase the use of simulation tools of increasing fidelity. For computational fluid dynamics (CFD), this can be moving from Reynolds-Averaged Navier Stokes (RANS) simulations to Large Eddy Simulation (LES) and eventually Direct Numerical Simulation (DNS). Thus, better methods of design space exploration are needed to prevent outpacing available computational or personnel resources during the course of design or analysis activities, even with efficient DoE approaches.

Use of an adaptive sampling (AS) scheme, which selects new points in the design space to explore based on all prior data collected and/or the metamodel used to model the data, is one avenue of improvement in this regard. AS works on the principle that sampling should be focused in either regions of high metamodel uncertainty or in regions where global maxima are found. By using an AS scheme, metamodel accuracy may be improved more rapidly relative to fixed sampling approaches for the same number of samples. Alternatively, AS may require fewer points to be selected relative to fixed sampling for some pre-defined level of model convergence.

The purpose of this study was to evaluate the use of a simplified AS methodology on an example problem of engineering interest. The response of interest was an engine performance parameter calculated using a lower fidelity cycle code. The metamodel used to approximate the system response was of least squares polynomial form with model variance quantified using a delta methodology. An AS scheme was used that determined the next best sample based on the maximum standard error in the design space. The performance of the adaptive sampling scheme was compared against a fixed sampling (FS) scheme, in this case Latin hypercube sampling (LHS) (McKay, Beckman, and Conover 1979).

METHODOLOGY

Metamodeling

Metamodeling is a technique by which a system response of a group of numerical or experimental simulations are approximated using a limited dataset over a defined design space. This technique is also known as surrogate modeling, reduced order modeling, and emulation; however, metamodeling is the preferred term here. The use of a metamodel is advantageous in design and analysis situations where rapid design space interrogation is required, for example in situations where repeated function evaluations of a high fidelity model may otherwise be required.

Advances have been made in the form of metamodels that can be used to approximate a system's behavior, for example kriging via Design and Analysis of Computer Experiments (DACE) (Sacks et al. 1989); however, a classical and common engineering approach to the modeling of a system response is to use a polynomial equation fit. Best practice dictates that a number of additional samples be used above the number of coefficients present in the model in order to produce a least squares fit representing the mean behavior of the response along with some modeling error. Thus, the final fitted form of the metamodel is

$$\hat{R}(\mathbf{X}) = [1, x_1, x_2, \ldots, x_1^2, x_2^2, \ldots, x_1 x_2, \ldots] \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_{11} \\ \hat{b}_{22} \\ \vdots \\ \hat{b}_{12} \\ \vdots \end{bmatrix} + \varepsilon = \mathbf{X}^\mathbf{T} \hat{\mathbf{b}} + \varepsilon, \tag{1}$$

where $\hat{R}$ is the modeled system response of interest, x_i is a design variable under the control of the analyst, $\hat{b}_i$ is a regression coefficient, and ε is a term reserved for the characterized metamodel error. In this example, a second order model with linear, quadratic, and interaction terms is shown, although higher order terms may be used to better approximate nonlinearities in the system response.

Metamodel Prediction Error

Quantifying the uncertainty in each simulation result was not within the scope of this study; however, the metamodel uncertainty with respect to the data used to fit it was quantified mathematically. The first step was to obtain the metamodel variance throughout the design space, which was calculated after determining both the sensitivity of the metamodel to the regression coefficients at a given point in the design space and the metamodel fit variance-covariance matrix ($\widehat{Cov}(\hat{\mathbf{b}})$), the latter of which is constant throughout the design space for a given fit. Equation (2) demonstrates this calculation using the delta method (Oehlert 1992),

$$\widehat{Var}(\hat{R}|\mathbf{X}) = \left[\frac{\partial \hat{R}(\hat{\mathbf{b}}, \mathbf{X})}{\partial \hat{b}} \right] \widehat{Cov}(\hat{\mathbf{b}}) \left[\frac{\partial \hat{R}(\hat{\mathbf{b}}, \mathbf{X})}{\partial \hat{b}} \right]^T, \tag{2}$$

where $\widehat{Var}(\hat{R}|\mathbf{X})$ is the variance of the metamodel at a given location in the design space.

Once the variance was obtained, it was transformed into the standard error for computing the confidence interval with the additional knowledge of the number of points, N, used to construct the model, as in eq. (3),

$$SE(\hat{R}|\mathbf{X}) = \sqrt{\frac{\widehat{Var}(\mathbf{X})}{N}}, \tag{3}$$

where $SE(\hat{R}|\mathbf{X})$ is the standard error of the metamodel at a given location in the design space.

Once the standard error was computed, model-form confidence interval bounds were calculated throughout the design space by multiplying the standard error by the Student's t value, which was obtained through knowledge of the number of samples and the degree of confidence required. Equation (4) is the error term in eq. (1), given by

$$\varepsilon = \pm t SE. \tag{4}$$

A 95% confidence interval approaches a t value of 1.96 as the number of samples approaches infinity. When the model prediction interval is alternatively desired, eq. (4) may be multiplied by the square root of the number of samples taken. Once a confidence or prediction interval is defined, eq. (4) can be added to or subtracted from the model prediction to yield the upper and lower limits of the model error.

Adaptive Sampling

Besides full factorial sampling, a common practice in engineering is to use fixed DoE sampling to evaluate simulation results efficiently, both in physical experimentation as well as computer modeling. While physical experimentation commonly uses classical DoE with replicates and blocking to better characterize random error, modeling and simulation approaches have tended to use space filling approaches, such as LHS, to better capture any high degree of nonlinearity in the interior of the system design space. Other examples of DoE include optimized designs (e.g., *I*-Optimal, *D*-Optimal) that attempt to lower the variance of the model given an assumed model-form. While such sampling methods generally produce satisfactory results in terms of model accuracy, they are purely *a priori* in the sense that they do not respond to the characteristics of the design space as more information is gained during the sampling process. Thus, AS methods are a potential path of process improvement to make data collection more efficient beyond what conventional DoE offers.

The AS scheme in this paper is based on the principle of searching for new points to sample that occur in the regions of the design space where the metamodel variance is greatest. This is analogous to adding points via maximizing the model mean squared error in kriging methods. This problem statement is given in a standard optimization form as

Maximize:

$$f(\mathbf{X}) = \widehat{Var}(\hat{R}|\mathbf{X}) + P(\mathbf{X}) \tag{5a}$$

Subject to:

$$x_{i,min} \leq x_i \leq x_{i,max}, \tag{5b}$$

where f is the objective function to be maximized, and P is a penalty function that can be used to modify the model variance space. For this study, $P = 0$, but an alternative functional definition can be to have P negative at high magnitude near other sampled points in the design space and low magnitude far away from other points. This would allow greater dispersion of points throughout the design space.

Because the objective function may have multiple local maxima, a heuristic approach was taken to sample the objective function throughout the design space densely and then pick the best point as the initial guess for a sequential quadratic programming algorithm (Kraft 1988). In adddition, because a penalty function was not used on point spacing in this study, the algorithm had the potential to return to points previously sampled. In such cases, a replicate of that point was added at the previously sampled response. This did not contribute any inefficiency to the method and still allowed the algorithm to subsequently choose new candidate points.

Evaluating Metamodel Adequacy

Completing metamodel adequacy checks allows the analyst to understand how accurate any given metamodel is in representing the underlying dataset of the system behavior. Quantitative and qualitative goodness checks were used as indicators of metamodel adequacy and include $R^2_{adjusted}$, $R^2_{predicted}$, actual versus predicted plots, residual versus predicted plots, and the distribution of studentized residuals. These quantities are defined as follows:

$R^2_{adjusted}$: The adjusted correlation coefficient is related to the correlation coefficient but also includes penalties to account for the number of modeling terms used, and will therefore always be less than the unadjusted correlation coefficient. Ideally, it should be greater than 0.99 for engineering metamodeling applications.

$R^2_{predicted}$: The predicted correlation coefficient is a cross-validation metric that removes one fit point from the metamodel at a time and refits the metamodel. It essentially validates the metamodel against points used to fit the model. Ideally it should be close to the adjusted correlation coefficient to provide confidence that the mean metamodel is relatively stable to removing and adding points.

Actual versus Predicted Plots: The actual versus predicted plots should contain points that run closely to the 1:1 line with low dispersion about the perfect fit line and have no fanning or otherwise contain any other nonlinear trends.

Residual versus Predicted Plots: The residual versus predicted plots should appear to have random scatter about the zero residual line with no observable trends that would indicate the metamodel is biased in any particular region of the design space.

Studentized Residuals: A key assumption to the least squares method is that the error is normally distributed about the mean prediction. Plotting a histogram of studentized residuals of the metamodel relative to the data fit points is a way to validate this assumption. Ideally, the residuals should have a normal distribution with zero mean and a standard deviation of one. The Shapiro-Wilk test (Shapiro and Wilk 1965) was used to determine the probability that the data do not adhere to a normal distribution.

EXAMPLE APPLICATION

The AS scheme described in the current paper was exercised in a physics-based modeling and simulation environment to predict the performance of a generic, non-vehicle-integrated, supersonic combustion ramjet (i.e., scramjet) cycle. A scramjet (as in fig. 1(a)) is an atmosphere-breathing engine designed to operate in the hypersonic regime of flight (i.e., $M \geq 5$, or greater than five times the speed of sound) in order to take advantage of inlet shockwaves to compress the flow rather than using turbomachinery (e.g., turbojets, as in fig. 1(b)). For the interested reader, a historical overview of international scramjet development is available (Curran 2001). A notional cutaway of a scramjet flowpath is shown in fig. 2, with the components of the flowpath labeled. For the purposes of this study, a propulsion cycle code with an equilibrium jump combustor analysis was used to estimate the performance response of the system throughout the design space. This allowed data to be collected quickly to evaluate the methodology in this study without using higher fidelity, longer run time codes (e.g., fully reacting CFD).
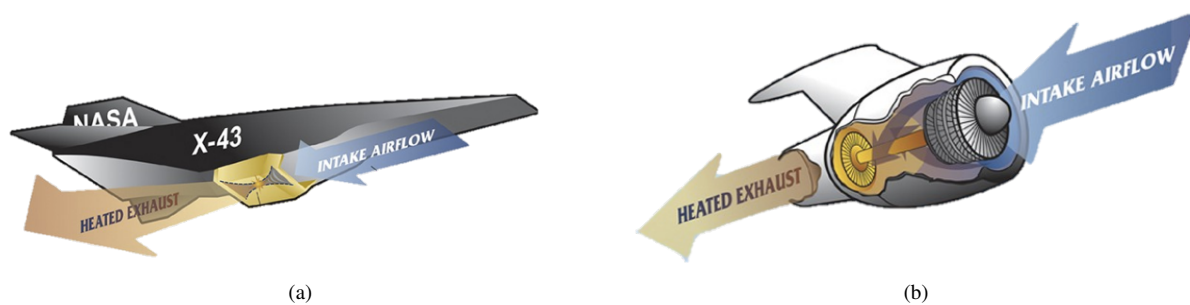

(a)   (b)

Figure 1. A comparison of a (a) vehicle-integrated scramjet engine and a (b) podded turbojet engine. Flow is from right to left.
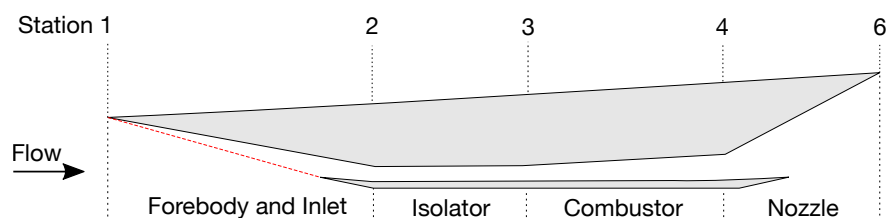

Figure 2. A notional schematic of a scramjet flowpath is shown. Flow is from left to right.

The response of interest was the maximum air specific impulse, which in this case is the uninstalled (i.e., non-vehicle integrated) thrust divided by the weight flow rate of air entering the system,

$$I_a = \frac{F_{un}}{\dot{W}_1}. \tag{6}$$

The air specific impulse is a nonlinear system response and a performance parameter for all air-breathing engines. It was calculated for a hydrogen-fueled system using the stream thrust method of Curran and Craig 1973. For this

study, the effect of Mach number ($6 \leq M \leq 12$) and inlet kinetic energy efficiency ($0.93 \leq \eta_{KE} \leq 0.97$) were the only parameters varied, with all other system parameters held to values that would yield idealized performance. The inlet kinetic energy efficiency reflects the entropy rise a supersonic inlet incurs due to the presence of shockwaves, and is defined via Smart 2012.

The metamodel form of the air specific impulse was defined in polynomial form to be

$$\hat{I}_a(\mathbf{X}) = [1, M, \eta_{KE}, M^2, \eta_{KE}^2, M\eta_{KE}, M^3, M^4] \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_{11} \\ \hat{b}_{22} \\ \hat{b}_{12} \\ \hat{b}_{111} \\ \hat{b}_{1111} \end{bmatrix} + \varepsilon = \mathbf{X^T}\hat{\mathbf{b}} + \varepsilon, \tag{7}$$

with higher order terms required due to the highly nonlinear nature of the air specific impulse when varying Mach number. This form of the metamodel equation was found to produce reasonable predictions in prior, unpublished analysis.

RESULTS

Both AS and FS schemes were applied to the same example problem to compare their performance. For both sampling cases, the maximum number of sampled points, N, was set to 20. The FS scheme was LHS with optimal spacing across the extent of the design space. The AS scheme initialized the samples with the corner and center points. Then, in order to calculate an initial least squares solution to the underdetermined metamodel, higher order metamodel terms were removed until a least squares fit was achieved. For the following samples, an additional term was added back into the metamodel until the full number of metamodel terms was recovered. Therefore, N values of 9 and above used the full metamodel form of eq. (7).

For the purposes of calculating the performance of each model, an additional 20 holdback points were sampled using LHS. Holdback points are not used during the metamodel fitting process; rather, these points were used to calculate the model response error (MRE) for each sampling scheme to understand the error characteristics of the metamodel created by each method. Figure 3 shows a visual summary of all the points sampled for each method, as well as the holdback points. Therefore, each method used a total of 40 samples, with 50% each being allocated to model fitting (via AS or FS) and to model validation.

Figures 4 and 5 show the final model response and standard error contours throughout the design space, respectively. For convenience, the sampled points are plotted. Both LHS and AS were able to fit metamodels that recovered the same general character of the design space, which has a peak in performance at maximum η_{KE} and Mach numbers between 8 and 9. The sampling structure for both schemes, however, were very different. The LHS scheme, as expected, was fully space filling with good dispersion of the data points. The AS scheme only had three points sampled on the design space interior, with the rest of the samples prioritized on the perimeter and corners of the design space. This yielded error characteristics for the AS scheme that had less variation relative to the LHS scheme, which had lows and highs in standard error exceeding that of the AS scheme. Also worth noting is that the AS scheme had four replicates (at the design space corners). Because those points didn't need to be rerun for the case of a deterministic computer simulation, the AS scheme required 20% fewer code evaluations relative to the LHS scheme.

Visualizations of model adequacy checks are shown in figs. 6 to 8. For the actual vs. predicted, residual vs. predicted, and studentized residual plots, no discernible bias is detectable. Furthermore, the studentized residuals both pass the Shapiro-Wilk test for normality with the normal distribution best fit overlaid on the histograms in fig. 8. Neither sampling scheme produced any outliers during the model fit process. The $R^2_{adjusted}$ values for AS and LHS were 0.96 and 0.97, respectively, and both had an $R^2_{predicted}$ value of 0.93. This indicates both models have converged to a similar level.
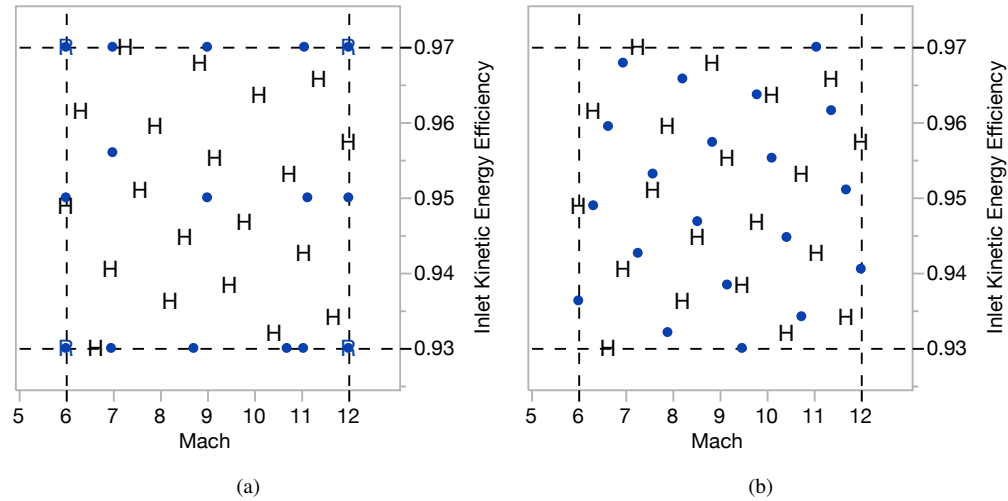
(a)

(b)

Figure 3. Multivariate plots for (a) AS and (b) LHS for *N* = 20. Circles are samples used for the purpose of model fitting. Holdback points are denoted by 'H' symbols and replicate sample points are denoted by 'R' symbols. Design space side constraints are shown with dotted lines.
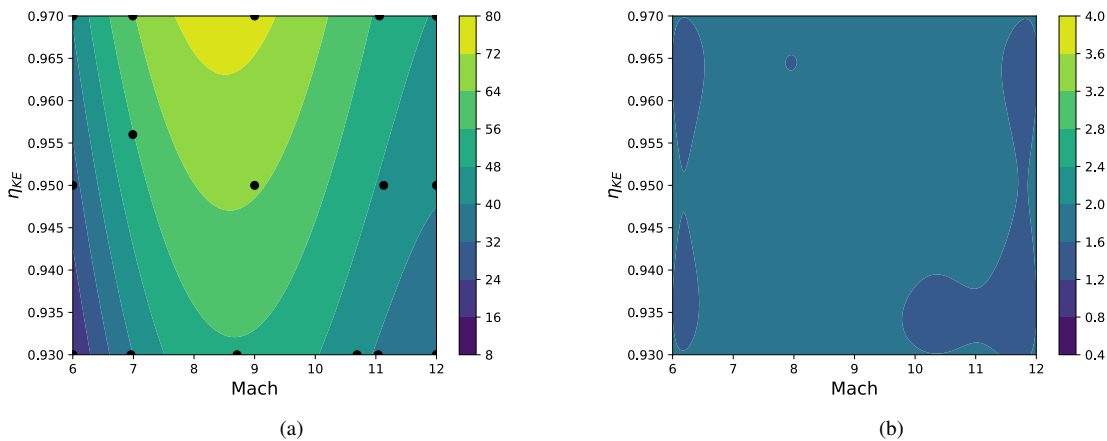

(a)

(b)

Figure 4. AS metamodel contours of (a) response and (b) standard error for *N* = 20, including four replicates. Sampled data points are denoted by the solid circles.

Next, the MRE for both sampling schemes were evaluated to test against holdback data not used to fit the metamodels. Figure 9 shows the histogram of the errors as well as the best fit distribution for each histogram: normal and generalized logarithmic for the AS and LHS metamodels, respectively. Table 1 summarizes the MRE errors at each quantile level for both the AS and LHS schemes. Overall, the AS MRE errors were more balanced and at each quantile level were competitive with or better than the LHS scheme. In addition, the min/max errors were lower than the LHS model, indicating a better prediction of points that were sampled but not used to fit the metamodels. Also noteworthy was that the LHS metamodel had four holdback points that violated the 95% prediction interval of the metamodel, while the AS metamodel had two violations. These results may be characteristic of the hypersonic propulsion application on which these sampling strategies have been applied. In such applications, metamodel performance on or near the design space boundaries is key, which may not be captured well with space filling schemes such as LHS.
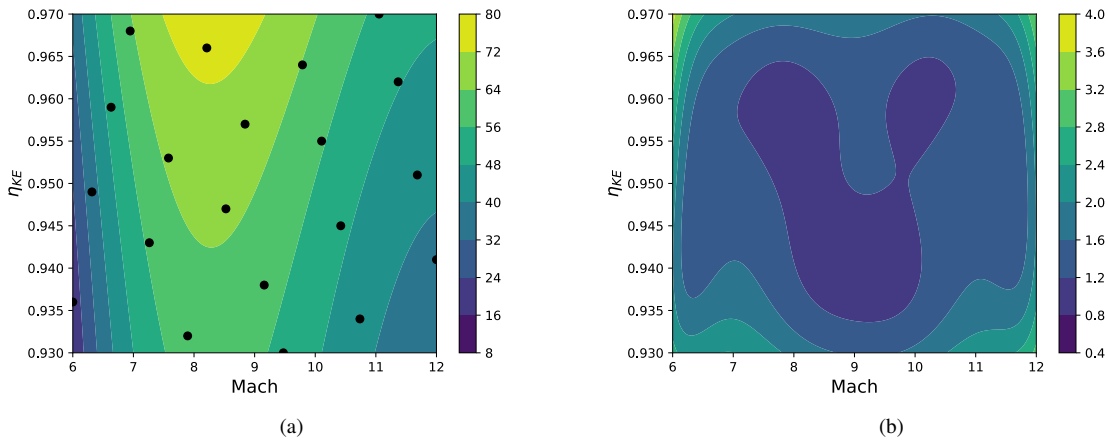
(a)   (b)

Figure 5. LHS metamodel contours of (a) response and (b) standard error for *N* = 20. Sampled data points are denoted by the solid circles.
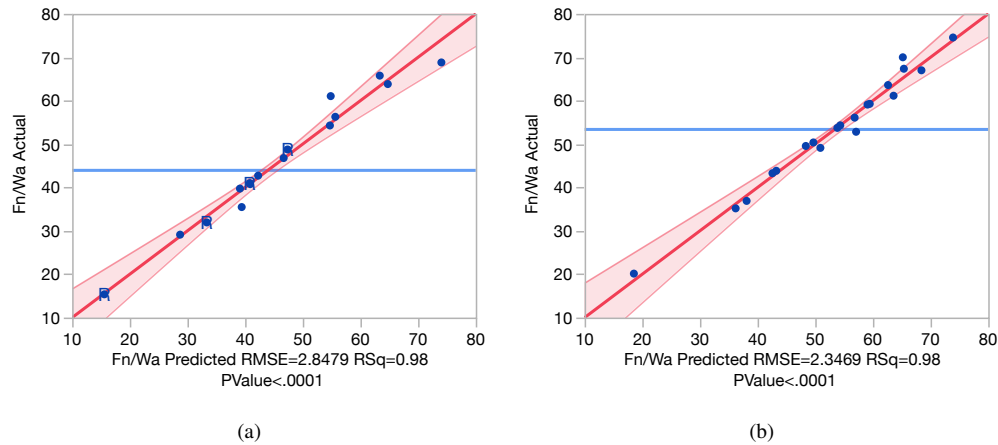

(a)   (b)

Figure 6. Actual vs. predicted plots for (a) AS and (b) LHS for *N* = 20.

Table 1. MRE errors by quantile for each method of sampling.

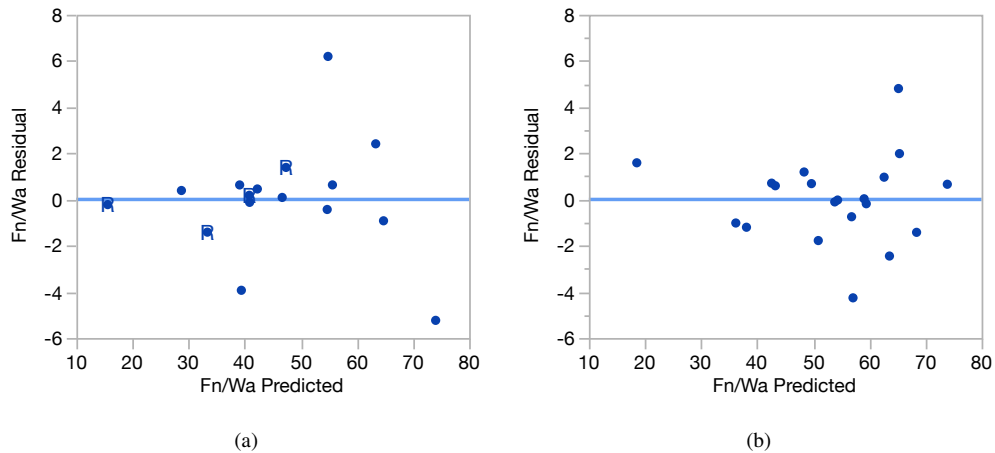| Quantile	LHS	AS
100%	17.83%	16.02%
99.5%	17.83%	16.02%
97.5%	17.83%	16.02%
90.0%	9.45%	10.09%
75.0%	4.14%	4.13%
50.0%	0.53%	-1.72%
25.0%	-3.57%	-5.17%
10.0%	-17.39%	-8.69%
2.5%	-39.24%	-15.98%
0.5%	-39.24%	-15.98%
0.0%	-39.24%	-15.98%

Figure 7. Residual vs. predicted plots for (a) AS and (b) LHS for *N* = 20.
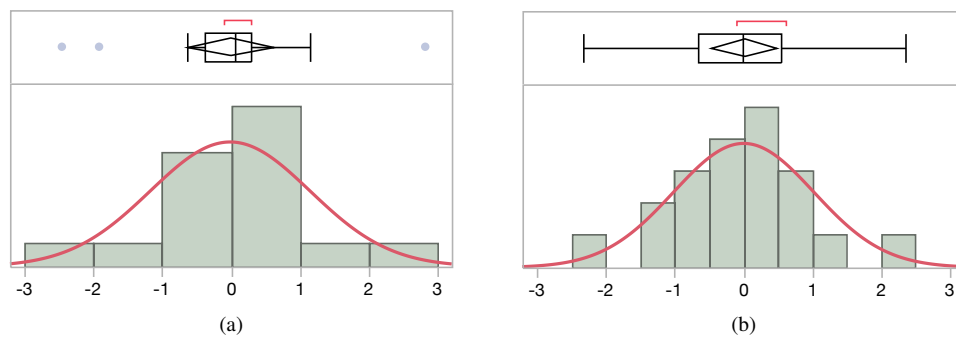

Figure 8. Studentized residual histograms for (a) AS and (b) LHS for *N* = 20.
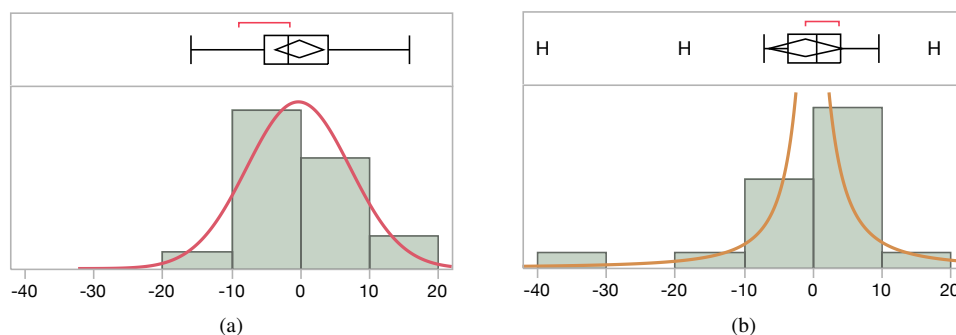

Figure 9. MRE histograms (by percent) for (a) AS and (b) LHS for *N* = 20.

SUMMARY AND CONCLUSIONS

Adaptive and fixed sampling schemes have been compared for an engineering problem of interest, including assessing metamodel adequacy for both sampling schemes and performance against holdback data. It has been shown that for 20% fewer function evaluations relative to LHS, metamodel performance was as competitive or better for AS relative to fixed sampling, even for a low dimensional problem such as the one tested in this study. Note that this performance may vary for other applications depending on problem nonlinearity, dimensionality, number of samples, and form of the underlying metamodel. However, this example shows the benefit of incorporating adaptive schemes into the

engineering process, especially early in the design and analysis process.

The focus of the current effort was to demonstrate the benefits of AS to a metamodeling method of common engineering application, least squares regression. While the impact of metamodel type (e.g., least squares regression versus kriging) was not a focus of this study, it is a worthy topic of comparison that is saved for future work. Worth noting is that kriging frequently finds application not only for adaptive sampling for metamodel refinement, but for optimization as well. Furthermore, augmentation to the present methodology to allow simultaneous, adaptive point selection with penalization of point selections near other points was not evaluated but is also saved for future work.

REFERENCES

Robinson, J. S. and J. G. Martin (2008). "An Overview of NASA's Integrated Design and Engineering Analysis (IDEA) Environment". In: *JANNAF 6th Modeling and Simulation/4th Liquid Propulsion/3rd Spacecraft Propulsion Joint Subcommittee Meeting*. Orlando, FL.

Myers, R. H. and D. C. Montgomery (2002). *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. 2nd Edition. Wiley-Interscience. ISBN: 9780471581000.

McClinton, C. R., S. M. Ferlemann, K. E. Rock, and P. G. Ferlemann (2002). "The Role of Formal Experiment Design in Hypersonic Flight System Technology Development". In: *40th AIAA Aerospace Sciences Meeting and Exhibit*. AIAA-2002-0543. Reno, NV. DOI: 10.2514/6.2002-543.

Bynum, M. D. and R. A. Baurle (2011). "A Design of Experiments Study for the HIFiRE Flight 2 Ground Test Computational Fluid Dynamics Results". In: *17th AIAA International Space Planes and Hypersonic Systems and Technologies Conference*. AIAA-2011-2203. San Francisco, CA. DOI: 10.2514/6.2011-2203.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). "A Comparison of Three Methods of Selecting Values of Input Variables in the Analysis of Output from a Computer Code". In: *Technometrics* 21.2, pp. 239–245. DOI: 10.2307/1268522.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). "Design and Analysis of Computer Experiments". In: *Statistical Science* 4.4, pp. 409–435. DOI: 10.1214/ss/1177012413.

Oehlert, G. W. (1992). "A Note on the Delta Method". In: *The American Statistician* 46.1, pp. 27–29. DOI: 10.2307/2684406.

Kraft, D. (1988). *A Software Package for Sequential Quadratic Programming*. Technical Report DFVLR-FB 88-28. Koln, Germany: DLR German Aerospace Center – Institute for Flight Mechanics.

Shapiro, S. and M. Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)". In: *Biometrika* 52.3/4, pp. 591–611. DOI: 10.2307/2333709.

Curran, E. T. (2001). "Scramjet Engines: The First Forty Years". In: *Journal of Propulsion and Power* 17.6, pp. 1138–1148. DOI: 10.2514/2.5875.

Curran, E. T. and R. R. Craig (1973). *The Use of Stream Thrust Concepts for the Approximate Evaluation of Hypersonic Ramjet Engine Performance*. Technical Report 73-38. AFRL.

Smart, M. K. (2012). "How Much Compression Should a Scramjet Inlet Do?" In: *AIAA Journal* 50.3, pp. 610–619. DOI: 10.2514/1.j051281.