Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1165

Published online 29 December 2010 in Wiley Online Library

Examining Improved Experimental Designs for Wind Tunnel Testing Using Monte Carlo Sampling Methods[‡]

Raymond R. Hill,[a*†] Derek A. Leggio,[b] Shay R. Capehart[c] and August G. Roesener[a]

Wind tunnels are used in the design and testing of a wide variety of systems and products. Wind tunnel test campaigns involve a large number of experimental data points, can take a long time to accomplish, and can consume tremendous resources. Design of Experiments is a systematic, statistically based approach to experimental design and analysis that has the potential to improve the efficiency and effectiveness of wind tunnel testing. In Defense Acquisition, wind tunnel testing of aircraft systems may require years of effort to fully characterize the system of interest. We employ data from a fairly large legacy wind tunnel test campaign and compare that data's corresponding response surface to the response surfaces derived from data generated using smaller, statistically motivated experimental design strategies. The comparison is accomplished using a Monte Carlo sampling methodology coupled with a statistical comparison of the system's estimated response surfaces. Initial results suggest a tremendous opportunity to reduce wind tunnel test efforts without losing test information. Published in 2010 by John Wiley & Sons, Ltd.

Keywords: experimental design; wind tunnel testing; Monte Carlo

1. Introduction

Wind tunnel testing has been a mainstay in aircraft design since the Wright Brother's experience at the turn of the 20th century. Wind tunnel testing is also valuable in the design and testing of a variety of products: automobiles, railway systems, building designs, camping tests, etc. Data from such testing are invaluable for achieving the required performance in such a variety of products.

Wind tunnel testing is a critical factor in the Developmental Test and Evaluation (DT&E) of United States Air Force (USAF) aeronautical systems including aircraft, spacecraft, and munitions, both those under initial development and those undergoing modification. In fact, essentially every system that interacts with the air medium has been or will be tested in the wind tunnel's controlled laboratory environment to assess the overall aerodynamic performance. While computational techniques, such as Computational Fluid Dynamics, have replaced a portion of wind tunnel testing, actual test results will always be required to validate these computational models.

Engineers routinely test design concepts, optimize configurations, or study stability and control (S&C) characteristics of new and modified aerospace systems, such as Small Diameter Bomb, F-22, and B-1B. A wind tunnel test may validate a modeling and simulation effort or serve as a precursor to open-air flight test. Ultimately, system program managers use wind tunnel output to support performance evaluation, trade-off studies, risk analysis, assessments of potential operational utility, and to demonstrate that the development process is complete and that the system meets specifications. These wind tunnel campaigns can often take a long time to complete with significant expenditure of resources.

The USAF and the Department of Defense are interested in improving the effectiveness of its wind tunnel test programs. The significant time and resources required for testing directly impacts the length of the system development cycle and the overall system cost. This fact is especially relevant as the USAF tries to replace its aging equipment in a politically turbulent and resource-constrained environment. Savings in the development schedule and/or budget are needed. The length of the

[a]Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433, U.S.A.
[b]Capability and Resource Integration, United States Strategic Command, Offut AFB, NE 68113, U.S.A.
[c]Department of Math and Statistics, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433, U.S.A.
*Correspondence to: Raymond R. Hill, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433, U.S.A.
†E-mail: raymond.hill@afit.edu

development schedule devoted to ground test is a focus of this research. In general, current high-performance military aircraft development programs require up to 3700 wind tunnel test hours in the conceptual design phase and up to 18 500 h in the development/validation phase[1]. The Arnold Engineering and Development Center (AEDC), a world-class flight simulation test facility, has sought 'reductions of up to 75% in costs and cycle time in developing and fielding new weapons systems'[2].

Conventional wind tunnel tests are based on an exhaustive data collection method that typically involves varying one independent variable (factor) across its entire range. This is often referred to as a sweep across that factor. This popular method is generally referred to as one-factor-at-a-time (OFAT) testing[3]. The OFAT approach does a good job of saturating the experimental design space with required design points. The OFAT method is, however, ineffective in terms of cost and effort as well as in controlling experimental uncertainty, detecting variable interactions, and producing minimum variance predictions. Nevertheless, it has been widely used since the time of the Wright Brothers despite great advances in experimental design addressing OFAT shortfalls.

In this research we examine whether systematically designed experiments can lead to a reduction in the overall required test effort without significantly degrading the amount of information generated during the test. Our specific focus is whether statistically designed experiments can reduce the number of wind tunnel design points without degrading the information generated and used to characterize the system performance. In general, wind tunnel testers are adverse to missing important system performance information and thus look for test campaigns, and the associated designs that saturate the experimental design space. In our focus, we examine whether the same response surfaces resulting from the current OFAT-based approach using saturated design regions can be attained with smaller, more sparse designs emanating from statistically based experimental design approaches.

To accomplish our research objective, we conduct a Monte Carlo experiment comparing the information obtained from a traditional wind tunnel campaign with information obtained via systematically defined experimental design strategies. To compare the information content obtained via either strategy, the resulting empirical surfaces are compared using both graphical and statistical methods.

2. Background

In a conventional OFAT wind tunnel test one independent factor is varied at a time. The factor of interest varies sequentially, usually in a single direction, through its selected range while the other factors are held constant. This is equivalently termed a sweep or polar. These incremental changes are often uniform in size. For instance, in aeronautical systems testing, angle of attack (AoA) polars may consist of two degree steps in the cruise region and one degree steps in the stall region[4]. The outputs of interest (responses) are recorded at each combination of factor settings (treatment). To illustrate, a common force and moment test may study Mach number, Reynolds number, AoA, sideslip, roll angle, and control surface positions as factors. Responses of interest would be lift, normal, and drag forces as well as the coefficients of the pitch, yaw, and roll moments.

An alternative to OFAT is to hold several factors constant while varying some select number of the remaining factors. Although this strategy may vary for several factors at once, it is also called OFAT due to the sequential setting of factor levels in each test run. Barlow *et al.*[4] provide a notional test program for an aircraft configuration evaluation using this OFAT wind tunnel experimental design approach.

OFAT tests contribute directly to the productivity goals of modern test facilities. Test facility productivity is often measured by the quantity of raw data collected with the allocated resources. Sequentially ordered factor level changes maximize the data acquisition rate, making OFAT tests highly efficient in terms of the total data volume. Tracking metrics such as 'cost per polar', 'cost per data point', 'polars per hour', and 'data points per facility per year' are common[5]. As a result, improvement in the wind tunnel community is often synonymous with collecting larger data sets. This mode of operation can be counterproductive to system development cost reduction goals and should be changed to a model for efficiently collecting data for the test. This is especially true when collecting data that are subsequently used to characterize system performance response surfaces.

Despite the widespread usage, OFAT designs have major deficiencies from a statistical and modeling point of view. They may produce biased data sets in the presence of time-varying systematic error. Creeping changes to test conditions (nuisance factors) such as set-point error associated with each chosen factor level, measurement error due to instrumentation drift, temperature variations, and changes in flow angularity can occur over the duration of a test entry. These errors can become compounded with the true factor effects, biasing the results. Prominent is OFAT's inability to efficiently detect interactions between factors. OFAT tests featuring non-varying factors at a baseline level throughout the test are blind to interaction effects despite the knowledge that in actual system performance evaluation, knowledge of factor interactions is essential.

None will deny that more test data is preferred, and more is rarely rejected. However, the size of a generated data set should not be the primary measure of that data set's quality. The prevailing view that 'data is the ultimate product of the test' discounts a key use of the created data sets: building predictive models[1]. Large data sets are often unnecessary for the efficient modeling of a system's aerodynamic characteristics. Further effort to gain more data can waste time and resources that can be put to better use supporting other aspects of system development and test.

Design of Experiments (DOE) is a statistical planning process used to systematically and efficiently assign factor combinations for each design point to ensure that the collected data can be analyzed using appropriate statistical methods. In many industrial applications of DOE, a first- or second-order response model sufficiently describes the region defined by the individual factor ranges (inference space). Wind tunnel models tend to be considerably more complicated and require higher precision[6]. Often, an appropriate model will require polynomial terms of third order or higher[7]. One method to avoid these higher order, more complicated models is to partition the inference space so that a first- or second-order response model will sufficiently describe each subspace. Partitioning benefits the aeronautical researcher by providing different models for fundamentally different flight regimes. For example, the model needed for the stall-vulnerable regions (usually high AoA) may be completely different from those needed for a benign cruise region. Thus, partitioning allows the explicit use of subject matter expertise in the choice of experimental design strategy within each subspace. The disadvantage of this approach is that more data and models may be required to describe the inference region. In practice, the researcher should seek the largest subspace ranges possible that can be modeled with low-order, less complicated functions[6].

Partitioning the inference space can lead to a response surface that may contain discontinuities at subspace boundaries. The predicted response offered by the models on either side of a boundary may differ in the mean value and in the width of prediction intervals. DeLoach and Erickson[6] compare this situation to the occurrence of random variation at a replicated point and state that concern is only warranted if the difference in the predictions from different models is larger than the required precision for the response. They point out that techniques are available to smooth the boundaries, but that this improvement in the perceived problem comes at the expense of less precision in both models[6].

The choice of experimental design is often unique for each project. DeLoach and Erickson[6] used partitioned inference spaces and low-order (quadratic or less) response models in their wind tunnel testing. They used fewer design points in well-behaved (e.g. nearly linear) regions and more design points in areas of highly nonlinear behavior. Morelli and DeLoach[8] used a Central Composite Design (CCD) augmented with a third-order, D-optimal design for each subspace. Dowgwillo and DeLoach[9] consider six subspaces for AoA and angle of sideslip and they used CCDs in each subspace augmented with a third block of points to support a sixth-order model. Landman *et al.*[10] used a Face-Centered Design (FCD) in a low-speed force and moment test of an aircraft with a complex control surface configuration. The factors examined were AoA, angle of sideslip, power level, and six control surface deflections. In short, experimental designs have been used to benefit in wind tunnel experimentation, but these past efforts have not tried to quantify the benefit of using statistical experimental design over traditional test approaches. We explicitly seek quantification of the potential benefit of statistically based experimental design for wind tunnel testing.

3. Methodology

Our research hypothesis is that systematic experimental design strategies can provide data sets sufficient for aeronautical system performance response surface characterization. Testing this hypothesis requires comparing response surfaces from both traditional and modern experimental design approaches. The comparison involves determining functional equivalency of the surfaces. Our performance metric is the reduction in design points when moving from an OFAT approach to a statistically based experimental design approach.

AEDC provided a legacy data set from a production wind tunnel test for a supersonic, expendable, low-altitude target drone. The system, designated here as the S-XX, simulated a high-speed, anti-ship missile. The test setup featured a 21% scale model of the system with actuated fin deflection. The test facility utilized model support with automated pitch capability and an automated roll capability. The missile was tested in both the transonic and supersonic ranges to characterize the overall aerodynamic performance as a function of attitude, control surface deflection, and speed. The set contained approximately 9000 runs (individual design points).

The first step in the work was to obtain an accurate response surface of the S-XX system based on the legacy data set of 9000+ runs. Multiple linear regression was used to approximate the true relationship between the six input factors ($X_1, \ldots, X_6$) and the output response, Y, revealed by the data set. Regression models created for production wind tunnel data may have R^2 values in excess of 0.9999[5]. These models often contain terms higher than third order[7]. For the purpose of this work, a reduced level of accuracy was acceptable and the models were limited to a full quadratic model and pure cubic terms. These models were built within a partitioned inference space so that a low-order polynomial model would sufficiently describe each subspace. The partitions used correspond to the transonic and supersonic subspaces. The response models constructed for each subspace are assumed to represent the true performance response surface of the missile system. Each subspace model was constructed in a similar fashion using similar model terms.

The model building process was quite thorough. Actual input parameter values were compared to the defined design levels; all regression models are based on design space settings. All factors were examined individually to assess the appropriate functional form for that factor. Multi-collinearity diagnostics ensured the use of independent factors in the model. Residual analysis confirmed standard regression modeling assumptions. Finally, outlier analyses detected potentially influential points ($<$5% of the points), but subsequent analyses led to retention of all such points.

The models were then validated by partitioning the data set into two subsets, a training set and a validation set. A rigorous model fitting exercise was conducted on the training set. The validation sample was used to re-estimate the regression coefficients of the model built from the training data. Cross-validation of the models examined the estimated regression coefficients, their estimated standard error, the error sum of squares (SSE), the mean SSE (MSE), and R^2 for both models. These comparisons provided assurance for the modeling approach. After validation, both subspace models were re-estimated with the full data set. The full intent of this initial phase was to accurately model the legacy wind tunnel data. Full details of the model building process and the validation effort are found in Leggio[11].

The best way to examine the efficiency and effectiveness of experimental design strategies in wind tunnel testing is to conduct the wind tunnel tests with those strategies. Naturally, time and resources prevented actual wind tunnel experimentation based on

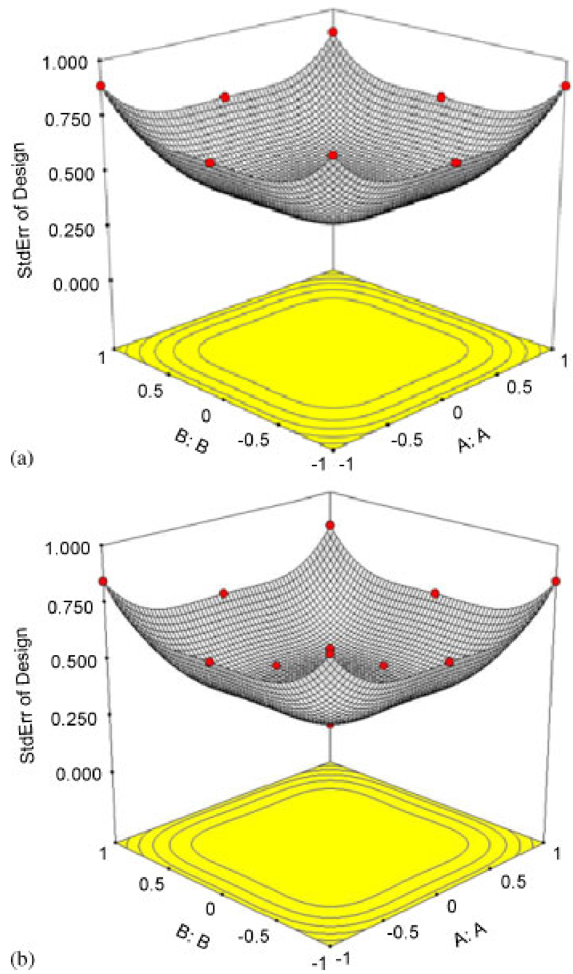

Figure 1. Standard error plot of FCD in (a) as compared to the standard error plot of NFCD in (b)

our experimental design strategies. Since the focus is assessing the equivalency of system response surfaces using differing test approaches, we are really examining whether experimental design strategies can replicate the information obtained via legacy methods at a fraction of the experimental costs (in terms of design points) associated with those legacy approaches that involve saturating the design space.

The empirical models fit to the actual wind tunnel data are assumed to constitute 'ground truth'. These models, with their high degree of fit and high values of R^2, represent what an experimenter would like to replicate with the data from a parsimonious, statistically based designed experiment. Thus, sampling from the ground truth models of the wind tunnel data serve as an adequate proxy for the wind tunnel experiments based on alternative experimental design strategies.

Our approach is to develop and assess alternative experimental design strategies. The designs selected, described below, are meant to be appropriate for the wind tunnel testing domain and within the scope of designs available to experimental design practitioners. Each of these strategies produce a schedule of experimental design points. These points are then used to sample from the ground truth model, with $N(0, \sigma^2)$ error added. The resulting data are then used to fit an empirical model based on data from the alternative experimental design strategy. These empirical models are then compared to the actual wind tunnel generated ground truth model using both graphical and statistical methods. A high degree of agreement between the models indicates some level of equivalence between the data collection strategies and correspondingly, low information loss when using the statistically based, experimental design strategy.

Selection of wind tunnel experimental designs is based on four key factors. First, the design must accommodate bounded factor ranges. The design must also accommodate higher order model estimation (for instance, estimation of both quadratic and cubic terms). Third, the design must be applicable to wind tunnel settings. Finally, the design must feature reasonable prediction variance properties through the experimental design region. Two alternative experimental design strategies were selected.

The first alternative design is the Nested Face-Centered Design (NFCD), introduced in Landman *et al.* [10] This is an FCD variant supporting quadratic models with pure cubic terms. The NFCD consists of an FCD inscribed within a larger FCD. The NFCD is designed for cuboidal design spaces, is particularly well suited for joining subspaces in a partitioned inference space, and fills the inference space in a fairly homogeneous fashion. Figure 1 shows the smooth changes in prediction variance from its basic FCD design (Figure 1(a)) to the NFCD (Figure 1(b)).

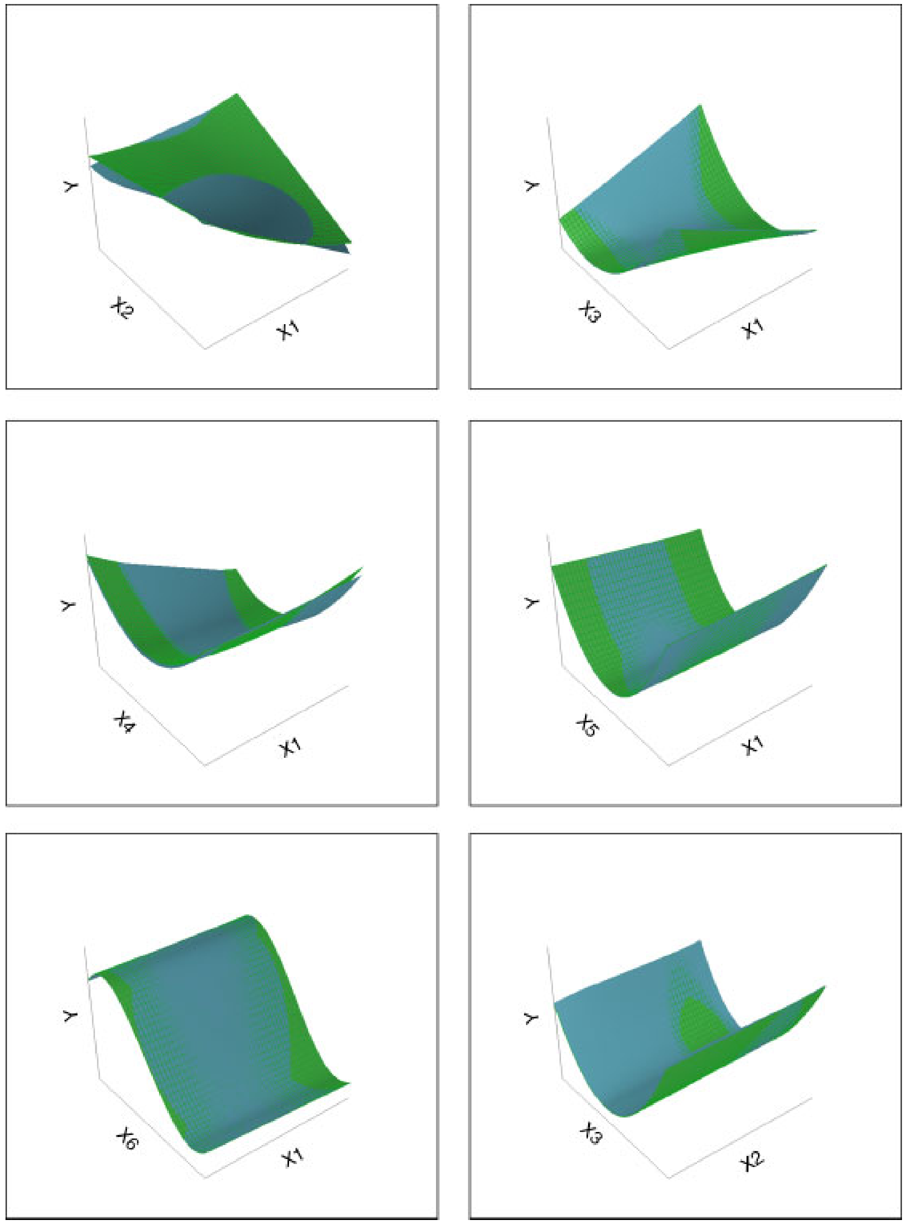

Figure 2. Comparison of surface GT transonic (green) and JD transonic (blue) (color in online version only)

The second alternative is the JMP I-optimal design. I-optimality minimizes the average variance of prediction over the region of the data[12]. Use of JMP allows for rapid, practical creation of the design in a cuboidal design space to fit the appropriate hypothetical model. The design may be blocked and the user may specify the number of factor levels so that a five-level design results. It is not limited by the subspace framework and is designed to provide the minimum average prediction variance across the entire subspace, including the boundary. The JMP software platform made this design particularly easy to construct and analyze.

Each alternative design provided a set of design points that were subsequently evaluated (or sampled) from the ground-truth surface based on the actual wind-tunnel data set. Random error was added to each evaluated point with the variance of the error chosen to emulate the error component from the ground-truth model. The data were then fit to appropriate multiple regression models *without using the knowledge pertaining to the ground truth model*. The same model building and model validation process previously described was used in the construction of these alternative models. The intent was to fit a model based on data from the reduced overall sample size of the alternative experimental designs and compare the surface information obtained via these alternative experimental designs and the ground truth surface obtained from the actual test data. We call the models GT, for the wind tunnel ground truth surface, ND, for the NFCD surface, and JD, for the JMP I-optimal design model surface.

4. Comparative methods and results

The comparison among the surfaces involved three comparison steps. The first was a graphical comparison involving subsurfaces of the surfaces. As the true response surface of the S-XX system in terms of the six input factors of interest cannot be displayed, we examined three-dimensional subspaces (2 factors and output). For each two-factor combination we overlay the surface plot of the GT response surface with the surface plot from an alternate experimental design (either ND or JD). Figures 2 and 3 provide sample results for transonic subspaces based on JD and the supersonic subspaces based on the ND, respectively. Other than some areas of minor disagreements, the reduced sample size experimental designs do quite well in matching the models obtained via the wind tunnel experimentation. These figures represent samples from the range of surfaces generated and are generally among the worst-case examples in terms of surface agreement.

The second method of comparison involved comparing the coefficients of the alternative models (ND and JD) with the coefficients from the ground truth model (GT). Table I presents the results for the transonic region. For each model, the coefficients are deemed equivalent at individual 95% confidence. Perusal of the coefficients support this equivalency. Table II presents the results for the supersonic region. Coefficients depicted in bold font in a column for an alternative model represent those coefficients that did not agree statistically with the corresponding coefficient from the ground truth model. Of the 21

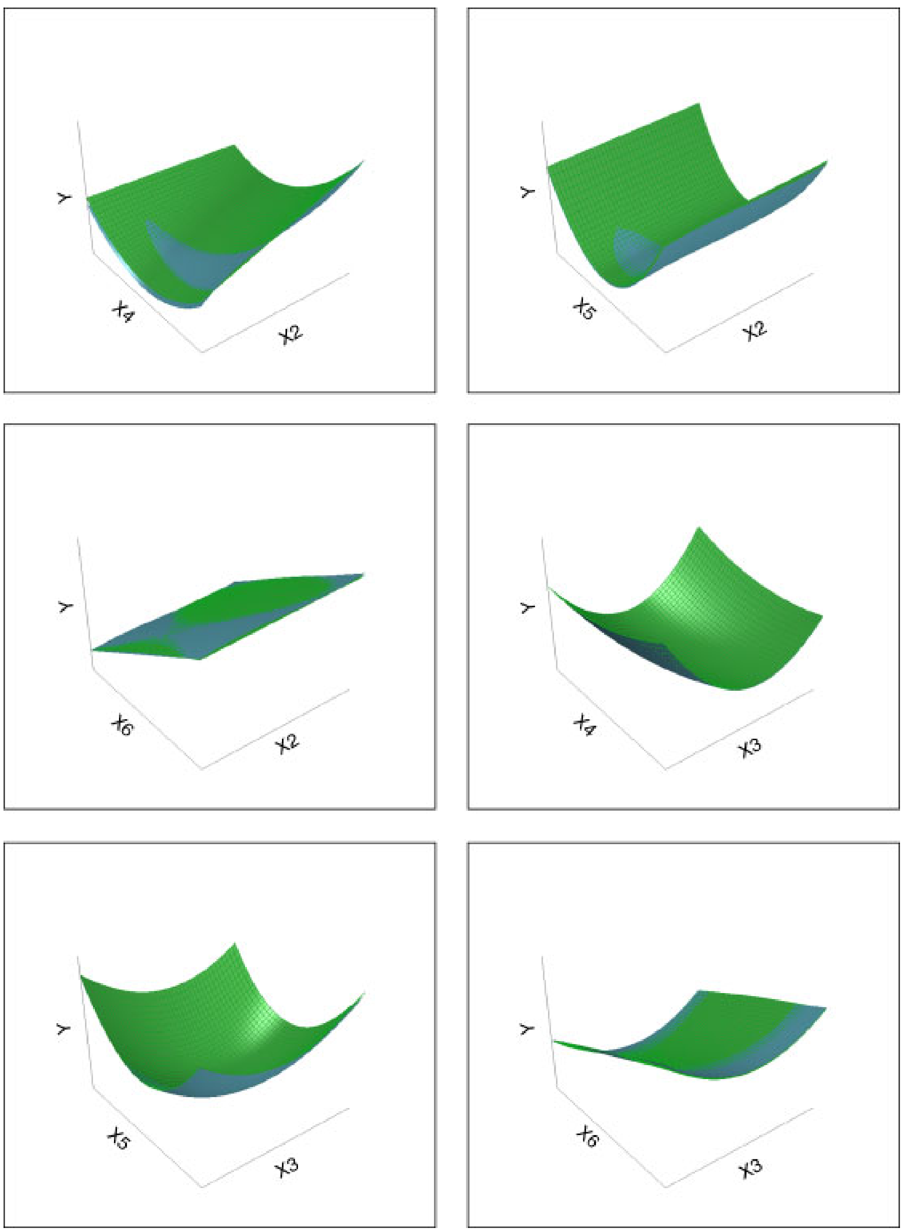

Figure 3. Comparison of surface GT supersonic (green) and ND supersonic (blue) (color in online version only)

Table I. Comparison of regression model coefficients for transonic subspace

	$Y_{T(GT)}$	$Y_{T(ND)}$	$Y_{T(JD)}$
Intercept	0.7276645	0.7287057	0.7277085
X1	−0.008916	−0.008691	−0.009471
X1ˆ2	0.0036692	0.0035258	0.0034736
X2	0.0052574	0.0046578	0.0050365
X3	−0.020997	−0.020727	−0.021482
X3ˆ2	0.1299117	0.1264269	0.1311797
X4	−0.010612	−0.010589	−0.010277
X4ˆ2	0.027947	0.0241082	0.0281005
X5	−0.035216	−0.03404	−0.03615
X5ˆ2	0.2434671	0.2471839	0.2422082
X6	0.6167071	0.6137196	0.6183015
X6ˆ2	−0.05499	−0.05539	−0.055593
X6ˆ3	−0.370656	−0.366468	−0.372917
X1*X2	0.0104301	0.0109318	0.0119782
X1*X3	0.0877043	0.0880601	0.0885876
X1*X4	−0.011519	−0.011949	−0.010273
X1*X5	−0.018356	−0.019986	−0.018282
X1*X6	0.0176079	0.0170921	0.0190207
X2*X4	−0.017622	−0.017955	−0.016913
X3*X4	0.0199821	0.0202817	0.0200294
X4*X5	−0.137442	−0.13679	−0.136495
X5*X6	−0.007419	−0.007566	−0.00738

Note: All sets of coefficients found consistent with each other and statistically equivalent.

Table II. Comparison of regression model coefficients for supersonic subspace

Term	$Y_{S(GT)}$	$Y_{SA(ND)}$	$Y_{SA(JD)}$
Intercept	−0.126954	−**0.121511**	−**0.123395**
X1	0.0591609	0.062423	0.0557311
X1ˆ2	0.009108	0.0049293	0.0101336
X1ˆ3	−0.037185	−**0.043455**	−0.035546
X2	0.006896	0.0070293	0.0072675
X3	−0.012877	−0.01436	−0.014203
X3ˆ2	0.0779585	0.0789256	0.0789877
X4	−0.006569	−0.006109	−0.007458
X4ˆ2	0.0174788	0.0198797	0.0184102
X5	−0.015172	−0.013992	−0.015272
X5ˆ2	0.1448347	0.1422188	0.1425919
X6	−0.075103	−**0.067652**	−**0.070627**
X6ˆ2	−0.004695	−0.000177	−0.004451
X1*X2	0.0065374	0.0072247	**0.0086192**
X1*X3	0.0561804	0.0556556	0.056513
X1*X4	−0.00621	−0.005874	−0.005736
X1*X5	−0.012867	−0.013858	−0.014099
X1*X6	0.0200284	0.018361	0.0191492
X2*X4	−0.008101	−0.009512	−0.008032
X3*X4	0.0159264	**0.0139634**	0.0156576
X3*X5	−0.01437	−0.013933	−0.013023
X3*X6	0.0037195	0.0042907	0.0034745

Note: Highlighted sets of coefficients statistically different.

non-intercept coefficients, 3 differ in ND and 2 differ in JD. Collectively, we find extremely strong agreement between the response surface models.

The third and final method of comparison involved statistically comparing the GT model with each of the ND and JD models. First, we suppose that the real unknown response surface of the S-XX system is well represented by GT. Each alternative

Published in 2010 by John Wiley & Sons, Ltd.

Qual. Reliab. Engng. Int. **2011**, 27 795–803

Table III. 95% CI for mean difference values between estimated subspace models and the GT subspace models

	Lower 95%	$\bar{D}$	Upper 95%
ND_T	0.00162	0.00169	0.00176
ND_S	0.00181	0.00166	0.00151
JD_T	0.00008	0.000133	0.00018
JD_S	0.00074	0.000814	0.00088

Subscripts are T for transonic, S for supersonic subspaces.

experimental design surface, ND and JD, represent the estimated surface such as obtained via testing. Let $\hat{Y}_{\text{GT}}$ and $\hat{Y}_{\text{AS}}$ be responses from the GT surface and then either alternative surface (AS representing either ND or JD), respectively.

An identical grid of data points was established denoted by $G=(x_{1J_1}, x_{2J_2}, x_{3J_3}, x_{4J_4}, x_{5J_5}, x_{6J_6})$, for $J_i=1,\ldots,n_i$, $i=1,\ldots,6$. Both surfaces were then sampled at each point in the grid, $g_k \in G$, for $k=1,\ldots,\prod_{i=1}^{6} n_i=K$, so that for all $k\in K$, $(\hat{Y}_{\text{GT},k}, \hat{Y}_{\text{AS},k})$ forms a sample of paired responses. The difference between the responses was defined as $D_k=\hat{Y}_{\text{GT},k}-\hat{Y}_{\text{AS},k}$, for $k=1,\ldots,k$. As the pairs of responses are independent and identically distributed, the D_k are independent and identically distributed with

$$\mu_D=E[D_k]=E[\hat{Y}_{\text{GT},k}]-E[\hat{Y}_{\text{AS},k}]$$

and

$$\begin{aligned}\sigma_D^2 &= V(D_k)=V(\hat{Y}_{\text{GT}})+V(\hat{Y}_{\text{AS}})-2\text{Cov}(\hat{Y}_{\text{GT},k}, \hat{Y}_{\text{AS},k})\\ &=\sigma_1^2+\sigma_2^2-2\rho\sigma_1^2\sigma_2^2.\end{aligned}$$

The unbiased estimator of μ_D is $\bar{D}$, where

$$E[\bar{D}]=\mu_D$$

and the unbiased estimator of $\sigma_{\bar{D}}^2$ is

$$\sigma_{\bar{D}}^2 V(\bar{D})=\frac{\sigma_D^2}{K}=\frac{1}{K}[\sigma_1^2+\sigma_2^2-2\rho\sigma_1^2\sigma_2^2].$$

As the individual responses $\hat{Y}_{\text{GT},k}$ and $\hat{Y}_{\text{AS},k}$ are approximately normally distributed for any given $k\in K$, the differences are normally distributed so that the statistic

$$t=\frac{\bar{D}-\mu_D}{\sigma_{\bar{D}}^2}$$

under the null hypothesis $\mu_D=0$ had a t distribution with $K-1$ degrees of freedom. This null hypothesis corresponds to the case where the surface estimated via the experimental design strategy is statistically equivalent to the surface determined by the actual wind tunnel test data.

The grid of 9450 points for each subspace was created. Table III contains the 95% confidence intervals resulting from the analysis. While none of the confidence intervals are statistically the same as zero, in a practical sense we claim that the surfaces in each of the subspaces are in fact sufficiently similar.

Thus, the response surfaces for this S-XX system are similar regardless of whether the full suite of 9000+ data points are used or a reduced suite of 900+ data points emanating from a statistically designed experiment are used. This is a 90% reduction in design points with no practical loss of information pertaining to the system response surface.

5. Summary and conclusions

Our methodology involved rigorously developing a surface accurately representing some 9000 data points from a legacy wind tunnel test campaign. We then treated this surface, labeled GT for ground truth, as an accurate representation of the true system, a system we labeled here as S-XX. With this surface we then consider two alternative experimental design strategies, an NFCD and an I-Optimal design from JMP. These design strategies were populated with responses obtained via a Monte Carlo sampling from GT. The data from the designs were used to generate surfaces ND, for the NFCD, and JD, for the I-Optimal design. A comparison of these surfaces with GT indicated little to no differences in either subspace for each of the surfaces.

This work represents an initial study. The complete wind tunnel test campaign studied involved nearly 9000 data points. The alternative experimental design strategies employed approximately 900 data points to cover the entire inference space consisting of the partition of the subspaces, transonic and supersonic. This is a 90% reduction in experimental effort. A reduction in experimental effort is quite easy to attain. However, a reduction in effort that is also accompanied by little to no loss in information is not so easy to attain. We established a case wherein the significant reduction in design effort comes with little to no reduction in the information generated.

This lack of difference among the surfaces leads us to an initial insight that we can in fact significantly reduce the overall effort involved in wind tunnel testing; in our case that reduction was 90% of the effort with *no* significant loss of data.

Naturally, future efforts abound. As an initial effort we cannot claim universal savings. However, we are looking at additional case studies. Our comparisons among surfaces is quite conservative; in future efforts we hope to exploit the subject matter expertise in the selection of experimental design strategies. Finally, we want to expand our look into those surfaces with higher levels of partitioning; this initial look involved very simple partitions of the inference space.

Acknowledgements

The authors thank the wonderful support of the Arnold Engineering and Development Center and Tom Best for the research topic and data provided to support this research.
Disclaimer: The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense or the US Government.

References

1. American Institute of Aeronautics & Astronautics (AIAA). AIAA Recommended Practice for Wind Tunnel Testing Part 2 (R-092-2-2003e): Practitioners Volume. AIAA: Reston, VA, 2003.
2. Peters WL, Lawrence WR, Mills ML, Milam WE. Cycle time reduction strategies and improvements in transonic testing in the AEDC Wind Tunnel 16T. *Proceedings of the 37th AIAA Aerospace Sciences Meeting and Exhibit. AIAA 1999-0179*. AIAA: Reston, VA, 1999.
3. Montgomery DC. *Design and Analysis of Experiments* (7th edn). Wiley: New York, 2009.
4. Barlow JB, Rae WH, Pope A. *Low-Speed Wind Tunnel Testing*. Wiley: New York, 1999.
5. DeLoach R. The modern design of experiments: A technical and marketing framework. *Proceedings of the 21st AIAA Advanced Measurement Technology and Ground Testing Conference. AIAA 2000-2691*. AIAA: Reston, VA, 2000.
6. DeLoach R, Erickson GE. Low-order response surface modeling of wind tunnel data over truncated inference subspaces. *Proceedings of the 41st Aerospace Sciences Meeting and Exhibit. AIAA 2003-456*. AIAA: Reston, VA, 2003.
7. Simpson J, Landman D. Low-speed wind tunnel testing via designed experiments: Challenges and ways forward. *Proceedings of the U.S. Air Force T&E Days. AIAA 2008-1664*. AIAA: Reston, VA, 2008.
8. Morelli EA, DeLoach R. Wind tunnel database development using modern experiment design and multivariate orthogonal functions. *Proceedings of the 41st Aerospace Sciences Meeting and Exhibit. AIAA 2003-653*. AIAA: Reston, VA, 2003.
9. Dowgwillo RM, DeLoach R. Using modern design of experiments to create a surface pressure database from a low speed wind tunnel test. *Proceedings of the 24th AIAA Aerodynamic Measurement Technology and Ground Testing Conference. AIAA 2004-2200*. AIAA: Reston, VA, 2004.
10. Landman D, Simpson J, Mariani R, Ortiz F, Britcher C. Hybrid design for aircraft wind-tunnel testing using response surface methodologies. *Journal of Aircraft* 2007; **44**:1214–1221.
11. Leggio DA. Examining methods to reduce wind tunnel test data requirements using the design of experiments. *MS Thesis*, AFIT/GOR/ENS/09-08, Air Force Institute of Technology, Wright-Patterson AFB, OH, 2009.
12. JMP. *Design of Experiments Guide, Release 7*. SAS Institute, Inc.: Cary, NC, 2007.

Authors' biographies

Raymond R. Hill is a Professor of Operations Research with the Department of Operational Sciences, Air Force Institute of Technology. His e-mail address is Raymond.Hill@afit.edu.

Derek A. Leggio is an Operations Research Analyst with the United States Air Force, US STRATCOM. His e-mail address is Derek.Leggio@us.af.mil.

Shay R. Capehart is an Assistant Professor of Statistics with the Department of Math and Statistics, Air Force Institute of Technology. His e-mail address is Shay.Capehart@afit.edu.

August G. Roesener is an Assistant Professor of Operations Research with the Department of Operational Sciences, Air Force Institute of Technology. His e-mail address is August.Roesener@afit.edu.