2012

Aerospace research through statistical engineering

Silvestrini, Rachel T.

Taylor & Francis

Quality Engineering, v.24, 2012, pp. 292-305
http://hdl.handle.net/10945/47555

Quality Engineering, 24:292–305, 2012
ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2012.641146

Taylor & Francis
Taylor & Francis Group

Aerospace Research through Statistical Engineering

Rachel T. Silvestrini[1], Peter A. Parker[2]

[1]Operations Research Department, Naval Postgraduate School, Monterey, California
[2]National Aeronautics and Space Administration, Hampton, Virginia

ABSTRACT The application of statistical engineering as described here is applied to fundamental research in support of hypersonic air breathing propulsion. An aerospace experiment that took place in a wind tunnel is the application of interest. One goal of the project was to accurately model the relationship between specific input and output parameters in the hypersonic combustion process within a controlled environment. Response surface methodology (RSM) was applied in a nontraditional manner to develop the response surface as the final deliverable rather than a tool to optimize a product or process.

KEYWORDS design of experiments, face-centered cube, repeat measures, replication, sample size

INTRODUCTION

This article presents the successful implementation of statistical engineering applied to aerodynamic research at NASA's Langley Research Center. The implementation of the statistical methodology was described in detail in a paper published by Johnson et al. (2009). In this article, we will revisit the supersonic combustion experiment but discuss the experiment to highlight the entire statistical engineering process rather than an emphasis on the statistical tools and methods.

Statistical engineering as described here is applied to fundamental research in support of hypersonic air breathing propulsion. The problem involved a high degree of complexity to accurately model the relationship between specific input and output parameters in the hypersonic combustion process within a controlled environment. The experiment was designed to characterize flow-field parameters as a function of the location, specifically axial and radial distance, within a coaxial free jet. The characterization experiment required the adaption and extension of response surface methodology (RSM) to study turbulent mixing properties with the goal to improve computational modeling.

The flow-field responses are unsteady (turbulent) due to instabilities; however, the experiment is designed such that the distribution of the flow parameters, mean and variance, are statistically stationary. Ultimately the researchers are interested in understanding the effect of turbulence through

Address correspondence to Peter A. Parker, NASA Langley Research Center, MS 238, Hampton, VA 23681, USA. E-mail: peter.a.parker@nasa.gov

a collection of statistics such as mean, variance, and covariance of parameters of interest throughout the entire jet engine flow field.

In this application, a statistical engineering approach was able to deliver an experimental approach to meet the specified objectives, and it clearly managed expectations about what could be obtained within the severe resource constraints prior to execution. As a result of this particular study and numerous other examples, the use of a statistical engineering approach is now becoming more widely accepted for these types of experiments. The impact is more efficient aerospace research that enables our knowledge to increase more quickly.

This article provides a case study of applying a statistical engineering approach to common class of problems encountered in aerospace research. We seek to highlight the application of statistical engineering and, in particular, the collaborative effort to establish a systematic method for solving complex problems of this nature. Moreover, we highlight efforts that NASA has made to improve the collaboration between statisticians and engineers/physicists. Aerospace research is a process of discovery, requiring a series of experiments that require efficient experimental design techniques. These techniques also often require the blending of many different statistical toolsets, combined with nonstatistical knowledge to fit the complexity and restrictions of the problem at hand. This article represents a successful illustration of the strategy for blending different statistical tools sets along with subject matter expertise that is applicable to many aerospace research projects.

BACKGROUND

Ground-based laboratory research facilities are used to simulate flight conditions to support a wide variety of aerospace research. In particular, recent research efforts include the study of supersonic and hypersonic flight and propulsion. Understanding the aerodynamic properties of supersonic and hypersonic flight is extremely important to the success of new aircraft that achieve speeds required by these flights. Barnstorff (2007) stated that these experiments help NASA engineers "address some of the challenges of hypersonic flight, a speed regime that's difficult to simulate and predict" (p. 11).

Experimentation in these types of facilities is complex and expensive. For these two reasons, it is important to develop statistically sound methods for developing and carrying out experiments that will be conducted in order to maximize the knowledge obtained. Developing statistically sound methods is a process that involves multiple subject-matter experts (SMEs) across multiple science and engineering disciplines. In the next section we introduce the statistical engineering approach that integrated computational fluid dynamics (CFD) modeling results, SME input, and several experimental design approaches to developing an overall experimental design strategy for a hypersonic wind tunnel test. An application of this methodology will then be illustrated, followed by a discussion.

METHODOLOGY

Statistical engineering is the terminology that describes the overall approach we used to develop a methodology that can be used to integrate CFD results and SME knowledge into experimental decisions as shown in Figure 1. As mentioned in the Background section, this problem is not unique to the application discussed in the next section; it is a reoccurring problem that could greatly benefit from a standardized approach.

Montgomery (2009) described statistical design of experiments as the "process of planning the experiment so that appropriate data will be collected and analyzed by statistical methods, resulting in valid and objective conclusions." This process includes a sequence of steps that start with preexperimental planning and end with conclusions and recommendations. For a detailed discussion on the process of designing and analyzing experiments, see Coleman and Montgomery (1993). Simpson et al. (2011)

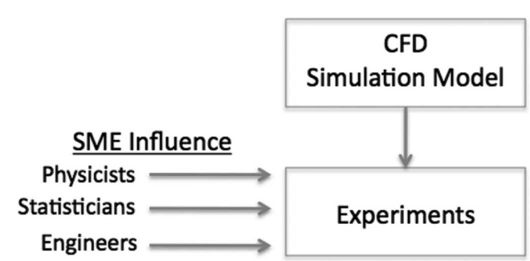


FIGURE 1 External influences used for the development of the experimental design used for testing. (Color figure available online.)

discussed the design of experiments (DOE) process as a circular cycle of experimentation. The steps in the circular experimentation process are summarized into four broad categories: plan, design, execute, and analyze. Figure 2 illustrates these steps and highlights the area of discussion, planning, and designing in this article. The integration of CFD model results and SME knowledge into an aerospace test requires an extensive iterative process in the planning and designing phase of experimental design and analysis. Though the focus of this article is on the plan and design steps, we will also briefly discuss the execute and analyze steps of the process through an application using a portion of CFD results.

Experimental designs are created in order to meet the goals of the problem as it is defined. The choice of design is based on a goal that typically falls into one of three categories: screening design, response surface methodology, or robust process design. Results from aerospace experiments are often utilized by many groups of people involved in the research process. Because of this, experiments often have a variety of goals that span multiple categories. Common statistical research tools and/or considerations that aerospace experiments and data analysis rely on include but are not limited to the following:

- Response surface methodology
- Robust process design (RPD)
- Complex statistical inference
- Data analysis for multiple responses
- Repeat measures
- Limitations of experiments due to physics constraints and/or equipment constraints

The theoretical foundation surrounding the tools and methods was used to create an integrated approach to wind tunnel experiments. The strategy used to integrate the tools was based on combining standard experimental design process in combination with SME involvement. Based on the needs of the project and the requests for information, it became evident that these tools were required throughout the process.

The first part of the experimental design process includes the planning phase as shown in Figure 2. In the methodology we suggest, we use the same preexperimental phases as given in Montgomery (2009), which include recognition of the statement of the problem, selection of the response variables, and choice of factors, levels, and ranges. In our experience, no aerodynamic experiment is focused on a single response variable, so we would like to emphasize the selection of multiple response variables. In this phase of the experimental planning, it is extremely important to have group meetings with all of the SMEs involved. This includes but is not limited to the physicists, statisticians, and engineers involved in both computational experiments and laboratory experiments.

The second part of the experimental design process covers the choice of experimental design. We have expanded this phase to include additional components beyond the goals of the experiment (which are identified in the planning phases under recognition and statement of the problem). The initial experimental design is modified by the use CFD results and modifications required by the physics and engineering concerns as depicted in Figure 3.

The overall process of planning and designing the experiment is presented in Figure 3. In the Application

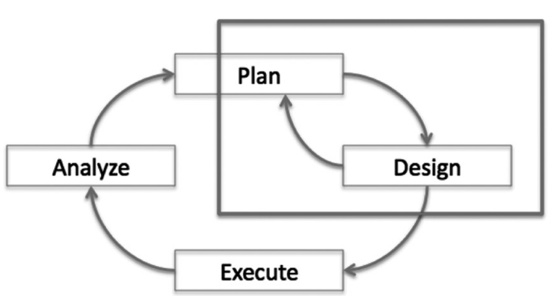


FIGURE 2 **Experimental design focus; the iterative process between the planning and designing phases in experimental design. (Color figure available online.)**

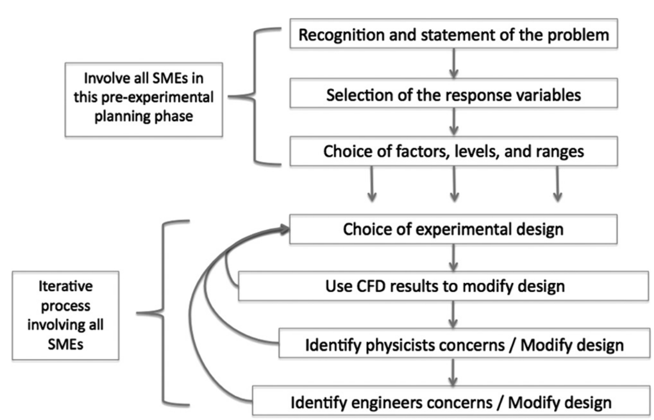


FIGURE 3 **Planning and designing phases of aerospace experimentation. (Color figure available online.)**

section, we identify the specific areas of research; for example, statistical inference, response surface methodology, and engineering modifications that influence the final choice of design.

AN APPLICATION OF AEROSPACE EXPERIMENTATION

This section is divided into three subsections: Planning (planning of the laboratory experiment), Designing (designing the experiment), and Executing and Analyzing (illustrating modeling approaches using CFD test data). An important facet of statistical engineering is that there is considerably more effort required to assist the SME to unequivocally define the objectives. In this example, we emphasize the solicitation and interpretation from the SME to ensure that we are solving the right problem.

Planning

Planning an experiment, as shown in Figure 3, includes all of the preexperimental design phases of the experimental design and analysis process. There are three main components to this phase: recognition of the statement of the problem, selection of the response variables, and choice of factors, levels, and ranges.

Recognition of the Statement of the Problem

The experimental design was created to ultimately characterize a number of flow-field parameters (response) as a function of the location within a coaxial free jet. From a physics and engineering perspective, this would allow the study and advanced research in turbulent mixing and computational modeling (for use in developing and validating the CFD). Ultimately the researchers involved in the project are interested in understanding the effect of turbulence through a collection of statistics. Essentially, the goals were to provide a comprehensive set of descriptive and inferential statistics and to characterize the response surface through the use of several different mathematical models.

At first glance, this goal seems to be RSM. However, the inferential statistics, complexity of the different responses, and the need to integrate CFD with the approach pulled this application out of the realm of RSM. This application involved a high degree of complexity, requiring more than one statistical technique to solve, and required both technical and nontechnical challenges, such as communication across a number of SMEs.

Choice of Response Variables

There are six main response variables of interest, shown in Table 1. In addition to these six response variables, it was of interest to characterize not only the mean of each response as a function of the input factors but also the variance and the covariance among pairs of responses. In addition to the six mean responses, there are six variance measurements (one for each response) and 15 (six choose two) covariance measurements. This makes a total of 27 responses of interest.

Choice of Factors, Levels, and Ranges

There are only two inputs of interest in the experiment, which are located within the open jet flame. The open jet flame is a three-dimensional area, but only two dimensions are modeled for the purpose of this experiment. An infrared image of the flame within an open jet can be seen in Figure 4, along with a picture of a cross section along the axis of the jet engine nozzle. (Note that the picture is compressed by a factor of 2 along the x-axis relative to the r-axis for the purpose of illustration.) Figure 4 also illustrates, for reference, the two input factors (x-axis and r-axis, both measured in millimeters) and their directionality within the open jet flame. It was assumed that the response would be symmetric about the z-axis (third dimension, going into the page), thus removing the need for a third input factor.

Designing

Developing an experimental design based on the unique goals and the complexity of aerodynamic

TABLE 1 Responses of interest in the wind tunnel tests

Temperature (K)
u-Axis velocity (m/s)
v-Axis velocity (m/s)
N_2 (%)
H_2 (%)
O_2 (%)

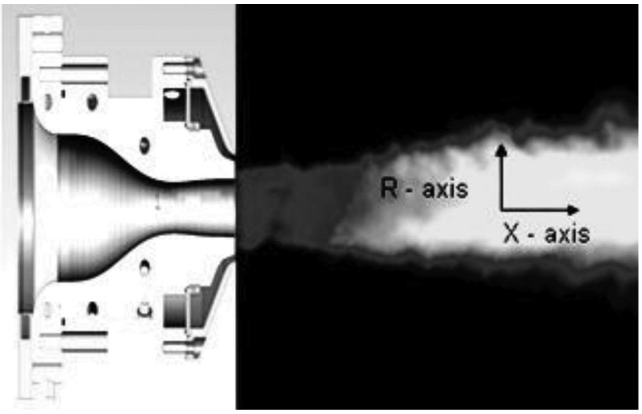


FIGURE 4 Diagram of the Mach 1.6 nozzle and infrared image of the open jet that issue from the nozzle.

research is considered a phase in itself. We treat the creation of the experimental design as a process that involves all of the SMEs interested in the problem. Though this phase is iterative, we discuss each of the four main components separately in this article. The components, as shown in Figure 3, are selection of the experimental design based on the statement of the problem (and goals of the experiment), use of the CFD results to modify experimental design choice, physicist SMEs' input/modification of the design, and engineers' input/modification of the design.

Note that using an experimental design approach for this problem is actually a unique application of experimental design in itself. This is because we usually think of experimental design as a way of controlling factors and studying their relationship (if any) with respect to a response. Our factors in this study are locations within an open jet flame. Thus, we are using experimental design as a sampling strategy.

Choice of Experimental Design

Employing design of experiments principles in hypersonic research has greatly improved the cost-effectiveness and information efficiency compared to previous methods. The choice of experimental design should always support the goals of the project and experiment. In the case of this application, the goals were quite broad and somewhat vague. There were so many SMEs involved in the experiments that a concise description incorporating all stakeholders' goals was challenging. As mentioned in the section "Recognition of the Statement of the Problem," the goals of the experimentation were to provide a comprehensive set of descriptive and inferential statistics and to characterize the response surface through the use of several different mathematical models.

The present experiment was an extension from a small-scale study to full-scale apparatus and was part of a much larger research effort in hypersonic air-breathing propulsion. As a result of this, there was prior knowledge available about several portions of the expected results. Prior knowledge about expected outcome of experiments can greatly aid the experimental design process. In particular, the prior knowledge provided the general nature of several of the response surfaces. This prior knowledge led to the establishment of several assumptions:

- Axis symmetric flow field
- Less drastic change in the x-axis direction than the r-axis direction
- Sharp points of inflection in the design space at the shear layer (see Figure 5)
- Constant gas flow rate during testing
- Unsteady flow-field parameters

The use of these assumptions allowed the experimental design to be developed in a more efficient manner. The axis symmetry assumption led to the removal of the rotational axis (as mentioned in the "Choice of Factors, Levels, and Ranges" section), meaning that the design was two-dimensional as opposed to three. The less drastic change in the x-axis direction than in the r-axis direction allowed for wider spacing of design points in the x-axis

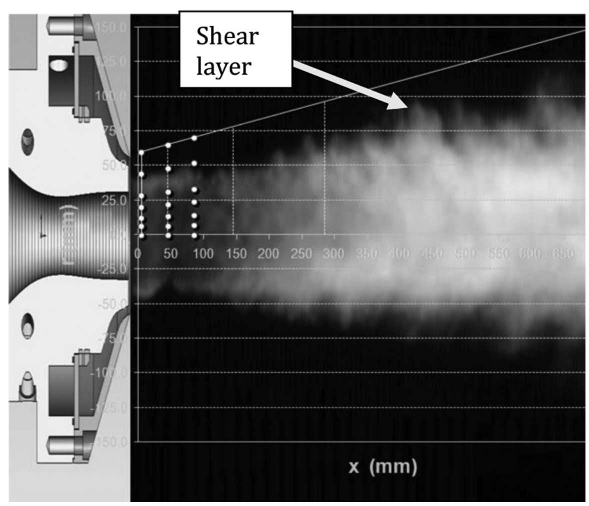


FIGURE 5 Angled–nested FCCD design based on peak angle with respect to the open jet nozzle.

direction and more densely packed points in the r-axis direction. Additional points at the shear layer location were found to be very important due to the sharp transitions of response variable outputs and the need for higher fidelity models around the shear layer. The flow-field parameters are expected to be unsteady (turbulent) due to instabilities; however, the experiment is designed such that the distribution of the flow parameters, mean and variance, are statistically stationary.

The data collected from the experiment will be used for the development of response surface models over different, sometimes overlapping, regions in the design space. Therefore, the precise estimation of single points of data, the ability to fit varying types of models, and the ability to fit models of both mean and variance were all of equal importance. Finally, the focus on the precision of the response surface for both mean and variance models of a particular variable was of interest. Modeling the relationship of the variance of a response variable for multiple random variables is not a topic that is often discussed or found in the literature. Vining and Schaub (1996) discussed designs for mean and variance functions of variables. Other useful references on this topic can be found in Myers et al. (2009). This statistical engineering application illustrates the use of a modified, partitioned classical experimental design in an innovative manner that accounts for the unusual design problem, design restrictions, and incorporates a robust method of systematically accounting for noise, or variability of the responses, inherent in the system.

In this case, RSM was applied in a nontraditional manner to develop the response surface as the final deliverable rather than a tool to optimize a product or process. In classical RSM, the goal is to approximate the response surface in order to specify optimal factor settings; in this usage of RSM we are not seeking to optimize the location within the flow field but rather to efficiently characterize the flow field to develop insights and to compare the experimental results with computational prediction. Though this may seem to be a relatively minor difference in our goal, it has a significant impact on the design. For example, we are interested in specifying a design that produces an absolute prediction variance, rather than a design with certain efficiency that essentially considers the reduction in prediction variance per point. In this case, we are interested in the unscaled prediction variance.

Another departure from classical RSM is our desire to develop a global response surface over the entire region of interest, or a partitioned quilt of response surfaces rather than a confirmed response surface in a local region of estimate optimum conditions, which is the classical RSM objective. Due to the complex and potentially discontinuous nature of the entire region of interest, the design region was partitioned to provide local second-order response surfaces in subregions of interest. This design approach also supports higher-order models over larger subregions and nonparametric modeling approaches.

Discussions about potential outcomes of the response variables, what types of models would be of interest to fit, and the design region lead to several options in the choice of experimental design that could be used. The following design types were considered:

- Optimal designs (D and I)
- Classical designs
- Space filling designs

Design D- and I-optimality were omitted as potential options early in the experimental design discussions due to uncertainty regarding the model over the entire space and the need to fit multiple models, perhaps in a piecewise manner, across the design region. Montgomery (2009) pointed out that optimal design criterion are of particular interest when the form of the model is known prior to experimentation. In this case, small-scale experimentation demonstrated very complex nonlinear models over large regions of the design space. Speculations about model forms over very small regions of the design space led to the possibility of piecewise second-order models.

The choice of a classical design often used in response surface methodology (Myers et al. 2009) is usually the central composite design (CCD). The CCD has the ability to fit up to a second-order model and consists of only nine design points in two factors. In two factors, the design has four corner runs, four axial runs, and a center run. In a spherical design space, the axial runs are set on the outside of the square region, so that the points lie along a circle surrounding the center point. Due to the near-rectangular design space, a face-centered central composite design (FCCD; also known as the face centered cube—FCC) was chosen over the CCD. In this design, the axial runs lie on the edge centers of a cuboidal space.

Space-filling designs are experimental designs that aim to fill the interior of a design region. The lower and upper bounds on the input variables define the design region. For a comprehensive discussion about different space filling designs see Santner et al. (2003) or Fang et al. (2006); also see Jones and Johnson (2009) for a graphical example of five different space-filling designs placed in a two-dimensional input design region.

The choice of space-filling designs is compelling for use in deterministic computer simulation models, where the output response does not have a stochastic component. The classical design option for response surface modeling was placed above the space-filling option because of the highly stochastic nature of the response variable, the experimenters' previous experience, and the need to fit models along radial traces (see x-axis locations; see Figure 4). Also, in designs such as the FCCD, near uniformity is attained, making it an attractive choice, especially if the need for a nonparametric model is foreseen.

Use of CFD to Modify Experimental Design

The FCCD design was chosen as the initial experimental design choice. The FCCD was chosen for several reasons, including the fitting of second-order models and its near uniformity. We mean that the FCCD has points spread equally throughout the region of design space, and this property is ideal for the fitting of both parametric and nonparametric models. The FCCD, though an excellent choice for RSM problems, was expected to have several drawbacks. The first sets of modifications to the design were created based on both infrared pictures of the small-scale open jet experiment and the CFD output results. The CFD model was a simulation model developed to mimic real aerodynamic properties of hypersonic flight.

Though a traditional FCCD (classical square shape) was used to start with, it was determined that the design needed to be set at an angle to the nozzle exit of the jet so that each of the points was on lines of a particular slope as opposed to parallel lines on the horizontal. This created a truncated cone shape instead of the traditional square. The reason for this modification of the classic design was due to the shape of the design region and the knowledge of the turbulent flow patterns throughout the open jet flame. The FCCD was set at angles corresponding to patterned turbulent flows exiting the nozzle head. See Figure 5 for the truncated cone shaped design region. Figure 5 also illustrates where the shear layer is located. From an experimental design perspective, the design region is a truncated cone shape with further restricted regions due to the data collection apparatus and its mobility limitations. In essence, this design approach mapped a nonrectangular truncated cone shape into a coded rectangular region where a FCCD was specified.

In addition to the modified shape of the FCCD, it was determined that a single FCCD would not provide enough degrees of freedom to fit the complex response expected in the r-axis direction because of the need for a much higher fidelity model over the larger design regions. In addition, the ability to possibly fit piecewise polynomial models in smaller design regions was desired. Therefore, it was decided that stacked FCCD designs would be implemented. Note that we could have chosen a single design with more levels (for example, a factorial with nine levels in each factor), which essentially would result in the same design we created. But because we wanted to visualize the piecewise polynomial models, so we chose to design and think about the factor space in terms of layered FCCDs.

Three FCCD design were stacked on top of the other with overlapping top and bottom points. Figure 5 illustrates three stacked FCCDs, which results in seven design points along each of three radial traces (r-axis lines at a given x-axis location). If the FCCDs were not overlapped, three FCCD design would create nine points along each of the three the radial traces; therefore, a saving of six design points was realized. Therefore, a stacked FCCD approach was used to fill the design space. In essence, the design consisted of a stacked and nested FCCD, which provides the largest preexperimental flexibility for modeling options over larger subregions or employing higher-order models. In contrast to a purely space-filling design, we incorporated all of the SME prior knowledge based on computational simulation and previous small-scale experiments to efficiently sample the entire region of interest.

The modified FCCD was attractive for two reasons: not only did it meet the ability to match patterns of the turbulent flow but it was also consistent with the shape of the flame that results after the combustion takes place.

The region of experimentation spanned a large distance in the *x*-axis direction (relative to the *r*-axis direction). Three FCCD designs as shown in Figure 5 were only enough to capture a small part of the design region just after the nozzle exit. One of the initial modeling assumptions listed was that there would be less drastic changes in the flow-field parameters, or response variables, in the *x*-axis direction. This assumption paired with the design space led to three main design regions of interest. The stacked FCCDs were therefore repeated in three separate design regions. The edges of each design region were overlapped so that savings on design point locations could be realized. A single region contained 21 design points, seen in Figure 5, so three regions would have a total of 63 design points, but because of the overlapping among the three regions of nested FCCDs only 49 design points were needed to cover the full region of interest. The adaptive scaling of the designs along the factor space based on prior knowledge dramatically reduced the number of design points while maintaining the ability to capture the behavior of the flow field.

Another modeling assumption mentioned was the sharp inflection points in the response variables in the radial axis direction due to steep gradients of variability in turbulent flow. At these points, even higher fidelity models were required. In the open jet experiment, nonconstant variance was expected, but no models of its form were known prior to experimentation. Only the knowledge that there would be an increase in turbulence, translating to extra variability, along the line known as the shear was available. The need for more points in regions of higher variance, without the knowledge of the form of the variance structure, led to the addition of points along several radial traces (or fixed *x*-axis locations in points of high gradient areas as seen from previous CFD models). Figure 6 shows several of the *x*-axis design locations with added radial (*r*-axis) points in regions of high gradient. Ultimately, the choice was made to add nine extra design points in the first region of the design space. This resulted in the modified FCCD design containing 58 unique locations.

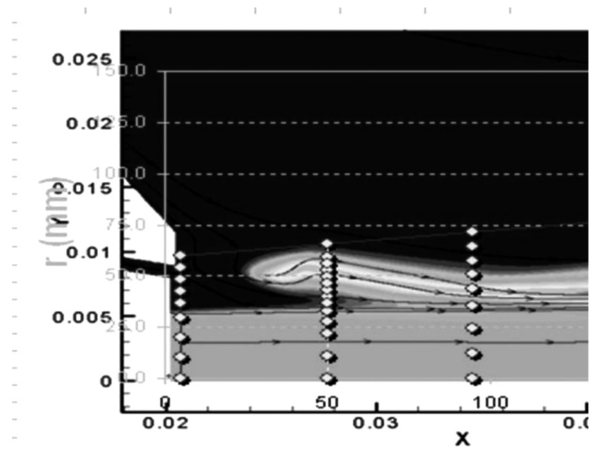


FIGURE 6 **Illustration of CFD results with added points in regions of high gradient.**

Physicists' Concerns and the Modification of the Experimental Design

Several of the SMEs involved in the wind tunnel experiments were interested in purely descriptive and inferential statistical measures of the responses. Of particular interest was the variability over time at a single location under homogenous experimental conditions. In order to achieve precise statistical metrics concerning the mean and variance of the responses, it was important to control the subsamples within a single design point. The idea of subsampling is different than replication in that the subsamples are repeat measurements taken by the measurement apparatus at a specific location prescribed by the design. Note that replications, which are visits back to that same design point location, are also important, especially in terms of the mathematical model. Replication is addressed in the next section on engineers' concerns.

In this section, we illustrate the development of the subsampling strategy by first introducing some basic statistical notation and then developing the method used in the experiment with a more sophisticated example and sensitivity analysis.

Let y_i be a single measurement at one location in the flow field, which we assume to be independent and identically distributed (iid) normal with a mean of μ_y and a variance σ_y^2. Let multiple measurements at a single location be denoted by a vector

$$\mathrm{y} = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_s]'$$

where s is the number of measurements at a design point. We refer to these readings as subsamples because multiple measurements at a single design point, without changing x-axis or r-axis (the inputs), are not replicate measurements but rather repeat measurements.

There are two main components that contribute to the variability among the measurements in $\mathbf{y}$. They are the measurement system noise (also referred to as precision of the measurement system), σ_m^2, and the variability in the flow media (turbulent flow), denoted by σ_t^2. Therefore, we can write

$$\sigma_{\mathrm{y}}^2 = \sigma_{\mathrm{m}}^2 + \sigma_{\mathrm{t}}^2 \qquad [1]$$

Note that we cannot separate the measurement system noise from the turbulence in a single experiment. Therefore, an estimate of σ_m^2 was obtained in a nonturbulent flow field. The estimation of σ_t^2 is intuitively restricted by σ_m^2. Using this notation, we can illustrate the estimate of the mean and variance of the response the vector $\mathbf{y}$ as, $\hat{\mu}_y = \bar{y}$ and $\hat{\sigma}_y^2 = S^2$, respectively. These statistics are both unbiased estimators and their variances are

$$\mathrm{Var}\left(\hat{\mu}_y\right) = \frac{\sigma_y^2}{s} \qquad [2]$$

$$\mathrm{Var}\left(\hat{\sigma}_y^2\right) = \frac{2\left(\sigma_y^2\right)^2}{s-1} = \frac{2\sigma_y^4}{s-1} \qquad [3]$$

We see that increasing the number of measurements, s, at a single design point decreases the variance of μ_y linearly and the standard deviation by $\sqrt{s}$. However, the variance of $\hat{\sigma}_y^2$ is a power of 2 larger than that of μ_y. Clearly, this implies the estimation of a variance requires more subsamples, a well-known result. See Casella and Berger (2002) for a discussion of Eq. [3].

Though Eqs. [2] and [3] describe the variance of flow-field parameters at a single design point location, we are also interested in identifying and quantifying the sources of variability over the entire flow field. To do this, consider taking multiple measurements at two different design points denoted by $\mathbf{y}_1$ and $\mathbf{y}_2$. The data collection protocol involves going to a design point (x_1, r_1) and collecting s_1 measurements and then proceeding to (x_2, r_2) and collecting s_2 measurements. We assume that there is a functional relationship between the distributional parameters and the location in the flow field. We can model the change in the mean response level (e.g., temperature) as a function of location within the flow field with a regression model:

$$\bar{y} = f(x, r) + \epsilon, \qquad [4]$$

where ϵ represents the residual regression error, and $f(x, r)$ is the functional form of the model. We assume ϵ to be iid normal with mean zero and a variance of σ_ϵ^2. The contributors to σ_ϵ^2 are defined to be experimental error ε and the lack-of-fit of the mathematical model, expressed as

$$\sigma_\epsilon^2 = \sigma_{PE}^2 + \sigma_{lof}^2 \qquad [5]$$

The experimental error, or pure error, is defined as the variability among the response measurements when all of the factors are precisely set to the same level, which includes the location in the flow field and the turbulence level. For example, we alternately set (x_1, r_1) and (x_2, r_2) collecting a vector of measurements at each location and compute the variability among the distributional parameters (because we are interested in estimating both the mean and variance). As we would expect, the estimated parameters will vary due to replication. This is independent of the functional relationship (mathematical model), because we have set all of the explanatory factors to identically the same level, within our ability to do so. In a response surface design, we include randomly allocated genuine replicates to estimate σ_{PE}^2.

To summarize the partitioning of the residual regression error, we have the σ_{PE}^2, which is a function of the two components of pure error identified earlier; measurement systems precision (σ_m^2) and the variability in the environment (turbulence, σ_t^2). Additionally, we have the σ_{lof}^2 component, which represents variability in the residual that could be explained by employing additional model terms. Thus, the partitioning of variance can be written as

$$\mathrm{Var}(\mu_y) = \sigma_m^2 + \sigma_t^2 + \sigma_{lof}^2 \qquad [6]$$

Using this equation we can see the different components of variance in the modeled response. The same expression can be written for modeling the mean variance of the response, where instead of $\mathrm{Var}(\bar{y})$ we have $\mathrm{Var}(\bar{S}^2)$, which is the average variability at a location within the flow field.

In this experimental design case study, one of the main design criteria was to specify the number of replications and subsamples required to achieve certain precision levels needed in order to estimate the mean and variance at a design point (locations within the flow field). We illustrated that the pure error is a function of the number of subsamples, s, taken to form one measurement, $\bar{y}$, used to fit the regression model. We can reduce the variance in the estimate of σ^2_{PE} by increasing the number of subsamples in the design; likewise, we can reduce the variance of the mean of y by increasing the replications because of the relationship

$$\mathrm{Var}(\bar{y}) = \frac{\sigma_y^2}{N} \quad [7]$$

where N represents the summation of all of the subsamples across the number of replications at a single design point i.

The number of replications plays a role in controlling the variance of the mean of the response variables, as seen in Eq. [7]. The number of subsamples directly impacts the ability to estimate the variance components on the flow parameters. Equation [3] demonstrates the number of measurements (subsamples) necessary to achieve a certain precision in the variance of a single output response variable.

In order to determine the subsample size required by the experimenters, it was important to have estimates of the variability components measurement and turbulence. Initial measures to acquire estimates were obtained through pilot runs of the system. Variability due to measurement error was realized with the open jet turned off and was assumed constant over the measurement range by the subject-matter experts. Measurements of turbulent flow variability were much harder to obtain, and expert knowledge was used to make reasonable estimates. Several estimates of the variability from the noise components were compared to study how the number of replications and subsamples would change based on the estimates.

Figure 7 is a plot of the number of subsamples needed to attain a specified precision in the turbulence measurement. Note that the experimenters preferred to focus on the precision of all measurements, whether mean or variance, in engineering units. Therefore, the variability in the turbulence measurements is expressed as the square root of the standard deviation, in this case, to result in units of degrees Kelvin. Different lines in the plot correspond to different estimates in the turbulent flow variability. Thus, Figure 7 demonstrates the sensitivity to the variance of the noise variables and its impact on how many subsamples are needed to achieve a particular precision level in the variance of an output response. For example, if the standard deviation of the noise variable contributed by turbulence was equal to 50 K (when studying the response variable of temperature), less than 10 subsamples would be needed in order to attain a precision in the standard deviation of temperature turbulence of 50 K. If the estimate of turbulent flow standard deviation was much higher, for example, 450 K, then well over 4,000 subsamples would be needed in order to attain the same precision of 50 K. This demonstrates the sensitivity of the effect of noise variables on the precision of variability of the response variables.

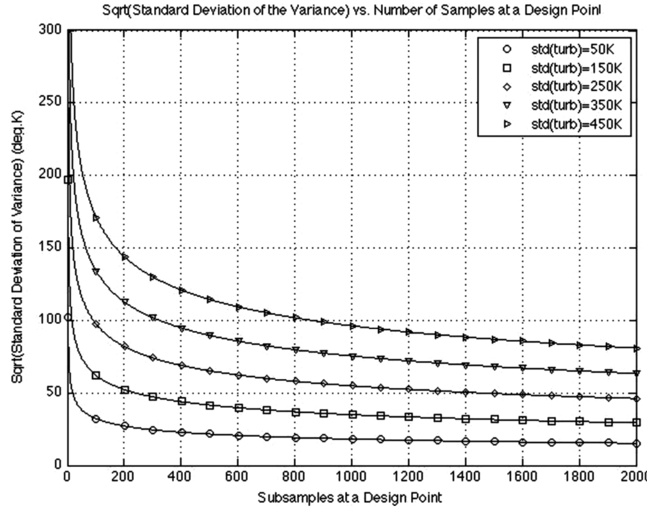


FIGURE 7 **Plot of standard deviation of variance of a response variable as a function of number of subsamples collected for several estimates of turbulent flow variability.**

Figure 7 demonstrates that achieving certain precisions levels for a variance response variable can be very expensive in terms of experimental effort, much more so than achieving precise estimates of the mean. After consulting with the experimenters, estimating the variance components, and determining the feasible amount of experimental effort, a decision was made to take 200 subsamples at each design point location and replicate a specific set of design points six times each. Subsampling 1,200 points (six times 200) represents a balance between the amount of experimental effort we could afford and the flat part of the curve in Figure 7, where adding more subsamples does not provide much additional precision. The balance between the precision of the estimates and the amount of time the jet could be operated during one design run was delicate.

Engineers' Concerns and the Modification of the Experimental Design

Three pillars of experimental design are replication, randomization, and blocking. Communication between the subject-matter experts' and the statisticians' understandings of what was possible allowed for the resulting choices in replication and blocking strategy. The open jet could only be run for approximately one minute at a time, due to the extreme heat levels and safety precautions. In this minute, only a total of 1,200 design points could be recorded if the data collection apparatus was kept in only one physical location. If the apparatus had to be moved to accommodate multiple points, points would be lost during the transition from one design point location to the other; the transition time was not instantaneous.

The final design featured 58 unique design points that specified physical locations across the experimental flow region (see Figure 8). At each design point, 200 subsamples were collected. Out of the 58 design points, 26 of them included replication that required moving away from the location in the flow field and returning to the set point. Twenty-one points were replicated six times and five points were replicated two times. Based on this replication strategy, the final design contained 168 runs at 58 unique locations. To further clarify the number of measurements obtained, at 21 locations 1,200 subsamples were collected, at 5 locations 400 subsamples were collected, and at the remaining 32 unreplicated points 200 subsamples were collected, resulting in 33,600 total measurements.

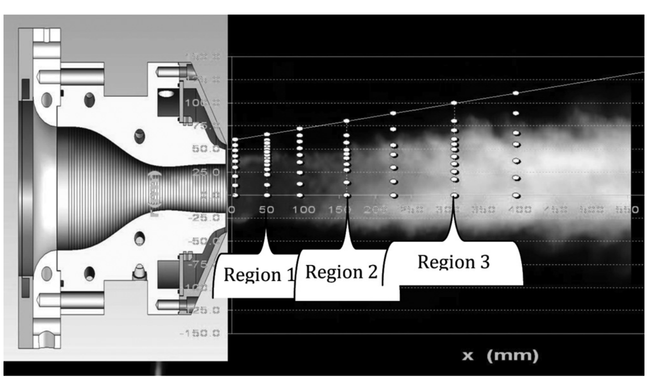


FIGURE 8 **Open jet with final design points depicted.**

Execute and Analyze—An Example

The purpose of this article is to highlight the statistical engineering approach required for planning and designing the experimental design strategy for a hypersonic wind tunnel test. We also present a brief portion on execution of the experimental design and analysis techniques. The data used for this article are based on CFD models that were created throughout the design process. Though these are not the data from the actual controlled wind tunnel testing, they highlight the approaches we use in the analysis of the actual data.

Execution—Implementing the Experimental Design Approach

CFD models were created to mimic the aerodynamic properties in the ground facility tests. The CFD data were used to influence the experimental design approach, as mentioned in the previous section. For this article, we use some of the results from the CFD models to illustrate the execute and analyze steps of the entire experimental design and analysis approach. The design points from region 1 are shown in Figure 9, which contains all of the simulated points

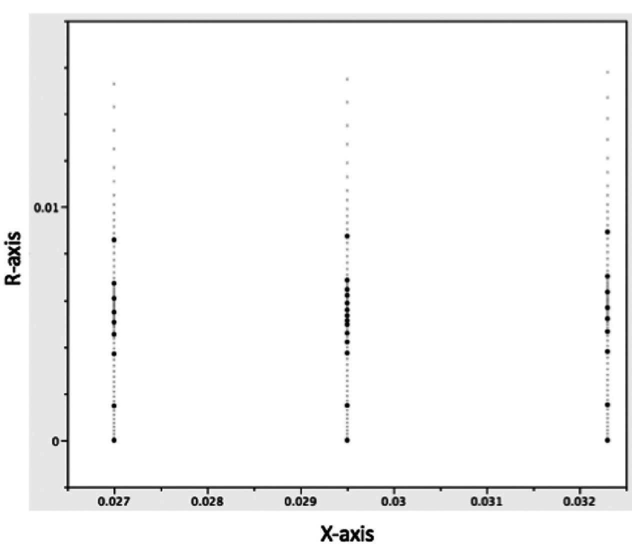


FIGURE 9 **Light grey points: all points created by CFD simulation model; black points: the experimental design points from region 1.**

along three radial traces of the experimental design. Note that there are many more design points simulated than points in the experimental design region. The CFD model, though computationally expensive to run, is not nearly as expensive or time consuming as running live experiments in the testing facility. In order to illustrate the analysis approach, as will be applied to the live data, we show the experimental design points from region 1 highlighted in Figure 9. The model fitting (conducted in the analysis portion of this section) will be done using only those points in the design region. The additional points will be used for cross-validation and model assessment.

The output data used for this analysis is temperature. The data from the 32 design points in region 1 are presented in Table 2.

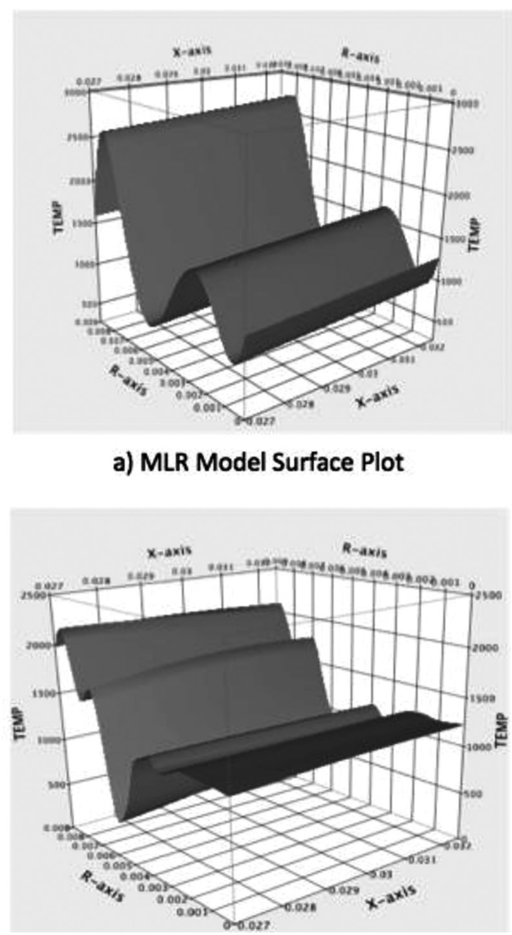
TABLE 2 Region 1 data from CDF model

x-Axis (m)	*r*-Axis (m)	Temperature (K)
0.027	0	1174.788
0.0295	0	1216.484
0.0323	0	1224.656
0.027	0.00147	1193.898
0.0295	0.00149	1193.538
0.0323	0.00152	1215.448
0.027	0.0037	1174.763
0.0295	0.00374	1168.031
0.0323	0.0038	1170.808
0.0295	0.00421	1157.463
0.027	0.00454	1137.278
0.0295	0.00459	1122.077
0.0323	0.00466	1044.493
0.0295	0.00496	844.721
0.027	0.00506	756.335
0.0295	0.00513	667.141
0.0323	0.00521	641.75
0.0295	0.00534	547.907
0.027	0.00549	456.941
0.0295	0.00558	522.849
0.0323	0.00568	711.347
0.0295	0.00588	628.252
0.027	0.00608	544.586
0.0295	0.00621	974.329
0.0323	0.00635	1212.878
0.0295	0.00646	1311.548
0.027	0.00672	1486.884
0.0295	0.00686	1755.907
0.0323	0.00703	1794.625
0.027	0.00859	2188.907
0.0295	0.00875	2154.386
0.0323	0.00893	2115.773

Analysis—Fitting Different Response Surface Models to the Data

Two response surfaces were chosen to test on this data set: a multiple linear regression (MLR) modeling technique and a Gaussian process (GP) modeling technique. Multiple linear regression modeling is a common choice for experimental design and RSM (see Montgomery 2009) model fitting. It is simple to do and easy to explain to SMEs who are not experts in statistics or mathematical model fitting. Gaussian process modeling has become popular as a modeling technique when used for fitting data from CDF models (an often cited paper is Sacks et al. [1989] and two books on the subject are Santner et al. [2003] and Fang et al. [2006]).

We treat the response of the data, temperature, as a realization of a stochastic random variable. Stepwise linear regression modeling was used to determine what factors, up through order 5, are significant in



b) GP Model Surface Plot

FIGURE 10 MLR and GP model surface plots.

TABLE 3 Fitted parameter estimates for the MLR model

Term	Estimate	Std. Error	t Ratio	Prob $>\|t\|$
Intercept	464.28	444.1448892	1.05	0.3059
x-Axis	25,063.24	14,677.11944	1.71	0.1001
r-Axis	−1,411,976.65	326,260.4993	−4.33	0.0002
r-Axis^2	1,664,543,975	290,695,319.3	5.73	<.0001
r-Axis^3	−6.02E+11	92,499,229,184	−6.51	<.0001
r-Axis^4	8.43E+13	1.21E+13	6.95	<.0001
r-Axis^5	−4.00E+15	5.58E+14	−7.17	<.0001

the x-axis and y-axis components of the inputs. The GP model is assumed to have the Gaussian correlation form; this is also the kriging model (Sacks et al. 1989). Plots of the response surfaces from the two models are shown in Figure 10.

Table 3 shows the fitted parameter estimates for the MLR and Table 4 shows the fitted parameter estimates for the GP model.

Using the extra points simulated by the CFD, the root mean square error (RMSE) was compared for each of the MLR and GP model fits. The RMSE for the MLR model was approximately 85,000, whereas the RMSE for the GP model was 212. In this instance, the GP model was much better than the MLR model. Though it is possible that the GP model could potentially overfit the data, we do not feel that this was the case here because of the low RMSE for the data that were held back. The MLR model does seem to be overfitting the data and this could potentially be avoided by not allowing such higher order terms. However, the removal of the higher order terms in the MLR results in extremely poor fits where the turbulent flow is high and where the temperature spikes in the region of interest. The GP model has been shown to be a very effective interpolator, especially for use in CFD output. Further investigation into modeling techniques for the actual data, including using alternative models and fitting to smaller portions (including piecewise linear regression) of the design region, will be undertaken.

TABLE 4 Fitted parameter estimates for the GP model

	Theta	Total sensitivity	Main effect
x-Axis	8,449.46	0.012759715	0.000768633
r-Axis	2,467,680.74	0.999231367	0.987240286

DISCUSSION

In this article we provide an expanded example of a statistical engineering approach to further the research into hypersonic propulsion. In many aerospace applications the use of design of experiments is the exception rather than the norm. There are numerous reasons, some of which are nontechnical, that hinder the routine application of statistical methods. One obvious reason is that there is no textbook solution available for our case study, and therefore a combination and extension of statistical tools were required. Someone exposed to an elementary introduction to design of experiments may conclude that it cannot be applied to our case. However, our case study shows that employing a statistical engineering approach that engineers the use of statistical sciences to achieve a better solution approach was successful.

Due to the problem's complexity, developing the experimental approach required close interaction among a multidisciplinary team that included statisticians, aerodynamic research scientists, measurement scientists, computational experts, and experimental facility operators. We emphasize that early involvement in clearly defining the research objectives in quantitative metrics that can be recognized when achieved was a significant departure from the typical role of the applied statistician, who may have been provided the factors and responses and simply asked to design an experiment in an independent effort. By employing a team approach, we sought to maximize the knowledge obtained from the experiment. A significant lesson learned in this example is that it takes the collaborative effort of a statistician and SME to solve complex problems and practice statistical engineering.

Once unequivocal metrics were obtained, fundamental statistical principles were applied and communicated to the team in a tutorial manner, which brought about buy-in from the nonstatisticians. A critical element of statistical engineering is having the statistician in a leading and/or full team member role during the problem solving, which often requires a much higher demand for education of other team members on statistical principles and thinking in addition to the ability to communicate those ideas into their subject-matter context. In this case, the communication of the number of subsamples necessary to estimate the variability within a prespecified precision was tantamount to the design of the experiment and reshaped the team's perspective on the execution of the experiment. Though this may seem obvious to an applied statistician, unless the subject-matter experts understood this important design criterion, they would have likely been opposed to the level of replication.

This example also illustrates the statistical engineering practice to link, adapt, combine, and/or extend existing statistical methods to solve a particular application. Combining fundamental statistical estimation principles with a nontraditional application of response surface methods enabled the design approach that supported parametric and nonparametric modeling methods. In particular, RSM principles, which were developed for product and process optimization, were applied to develop an efficient design strategy to create high-dimensional, global response surface models of both factor mean and variance. We note that this adaption naturally led to several promising statistical research areas noted in the article.

Overall, an application of statistical engineering was demonstrated to solve a large, complex problem that required a multidisciplinary team and synthesizing statistical principles and methods. We note also that our metric for success in applying statistical engineering was the multidisciplinary team embracing the proposed experimental approach as their own and not the statistician's approach.

ABOUT THE AUTHORS

Rachel T. Silvestrini is an assistant professor in the Operations Research Department at the Naval Postgraduate School. She received her Ph.D. and M.S. in industrial engineering from Arizona State University and her B.S. in industrial engineering from Northwestern University. Her research interests are in the design and analysis of both physical and computer experiments. She is a recipient of the Mary G. and Joseph Natrella Scholarship for excellence in statistics. She is a member of the American Society for Quality and the Institute for Operations Research and Management Sciences.

Peter A. Parker is a research scientist in the Aeronautics Systems Engineering Branch at the National Aeronautics and Space Administration's Langley Research Center. He holds a B.S. in engineering from Old Dominion University (1989), an M.S. in applied physics and computer science from Christopher Newport University (2000), and an M.S. (2003) and Ph.D. (2005) in statistics from Virginia Tech. He is a senior member of the American Society for Quality and the American Institute for Aeronautics and Astronautics and a member American Statistical Association. His research interests include experimental design and analysis, response surface methodology, statistical quality control and improvement, and the integration of computational and physical experimentation.

REFERENCES

Barnstorff, K. (2007). Wind tunnel tests help engineers go with the flow. Available at: http://www.nasa.gov/vision/earth/improvingflight/hybolt.html (accessed 6 September 2011).

Casella, G., Berger, R. L. (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

Coleman, D. E., Montgomery, D. C. (1993). A systematic approach to planning for a designed industrial experiment (with discussion). *Technometrics*, 35(1):1–27.

Fang, K. T., Li, R., Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Boca Raton, FL: Taylor & Francis Group.

Johnson, R. T., Parker, P. A., Montgomery, D. C., Cutler, A. D., Danehy, P. M., Rhew, R. D. (2009). Design strategies for response surface models for the study of supersonic combustion. *Quality and Reliability Engineering International*, 25:365–377.

Jones, B., Johnson, R. T. (2009). The design and analysis of the Gaussian process model. *Quality and Reliability Engineering International*, 25:515–524.

Montgomery, D. C. (2009). *Design and Analysis of Experiments*, 7th ed. New York: Wiley.

Myers, R. H., Montgomery, D. C., Anderson-Cook, C. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: Wiley.

Sacks, J., Welch, W. J., Mitchell, T. J., Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.

Santner, T. J., Williams, B. J., Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. New York: Springer-Verlag.

Simpson, J., Hutto, G., Sewell, A. (2011). Embedding DOE in military testing: One organization's roadmap. Paper read at the NASA Statistical Engineering Symposium, 3–5 May, Williamsburg, VA.

Vining, G. G., Schaub, D. (1996). Experimental designs for estimating both mean and variance functions. *Journal of Quality Technology*, 28:135–147.