AFTER FURTHER REVIEW: AN UPDATE ON MODELING AND DESIGN STRATEGIES FOR AGRICULTURAL DOSE-RESPONSE EXPERIMENTS

M. J. Frenzel

W. W. Stroup

E. T. Paparozzi

Recommended Citation

AFTER FURTHER REVIEW: AN UPDATE ON MODELING AND DESIGN STRATEGIES FOR AGRICULTURAL DOSE-RESPONSE EXPERIMENTS

M.J. Frenzel, Department of Statistics, University of Nebraska, Lincoln
W.W. Stroup, Department of Statistics, University of Nebraska, Lincoln
E.T. Paparozzi, Department of Agronomy and Horticulture, University of Nebraska, Lincoln

ABSTRACT: Research investigating dose-response relationships is common in agricultural science. This paper is an expansion on previous work by Guo, et al. (2006) motivated by plant nutrition research in horticulture. Plant response to level of nutrient applied is typically sigmoidal, i.e. no response at very low levels, observable response at mid-levels, point-of-diminishing returns and plateau at high levels. Plant scientists need accurate estimates of these response relationships for many reasons, including determining the lower threshold below which plants show deficiency symptoms and the point of diminishing returns, above which excessive doses are economically and environmentally costly. Guo et al. presented models and designs that address these requirements and a simulation study to assess and compare the small-sample behavior of these models and designs. This paper expands on that simulation study. In addition, a simulation study based procedure for exploring designs for experimental scenarios fitting this description is presented. This simulation study approach utilizes simulation based fit statistics in conjunction with various lack-of-fit plots to produce a design robust to multiple candidate models.

KEY WORDS: response surface, nonlinear regression, Central Composite Design, face centered cube design, Hoerl model, Gompertz model, Mitscherlich model, Logistic model

1. Introduction

A frequent focus of agricultural researchers is the characterization of the effect of two or more quantitative factors on some response. Common examples include the effect of pesticides on insect control or the effect of nutrient levels on plant growth. The latter example was the focus of a paper presented by Guo, Stroup, et al. in the 2006 conference proceedings. This paper begins by reiterating and expanding on the results presented by Guo, Stroup et al. while addressesing concerns and questions raised by the 2006 conference participants.

The work of Guo, et al. focused on research scenarios characterized by three main issues: the response variable of interest reacts nonlinearly to change in the levels of the quantitative factors, interaction among the factors is certainly present, and severe restrictions exist concerning the amount of time and experimental material available to the researchers. The working example used by Guo, et al. in the 2006 paper focused on the effect of varying levels of several elemental nutrients (specifically N, S, Fe, and Mn) on Poinsettia plants.

The nonlinear effect of nutrient levels on Poinsettia plants and the interaction between nutrient levels is depicted in Figure 1. The plants in this figure are arranged by increasing level of N on the left-to-right axis and increasing level of S along the front-to-back axis. At low levels of either N or S, the plants perform poorly, as indicated by the pale color of the plants. The poor

performance at a low level of one of the factors occurs irrespective of the level of the other factor, indicating an interaction between the two factors exists. The nonlinear pattern of response is evident based on the apparent diminishing returns with respect to both factors. As each factor reaches an appropriate level the plants appear much darker than at low levels of either factor. However, further increase in either factor does not increase the response, as plants exposed to medium levels of both N and S display the same color as plants exposed to high levels of both N and S. Thus, the characterization of nutrient levels on Poinsettia plants is, in essence, a dose response problem.

A common goal of studies addressing problems similar to the Poinsettia problem is to estimate the optimum dosage of the various factors, i.e. the combination of factor levels that produces the target response while using the minimum level of each factor possible to achieve this response. Clearly, any combination of factor levels that produces a response that is below the target is undesirable. In addition, excessive application of the factors beyond that required for the target response may be undesirable for economic or environmental reasons. The traditional approach to estimation of the optimum combination of factor levels is response surface methodology (RSM), which is described in many well respected texts (Box, Hunter, & Hunter, 2005; Myers & Montgomery, 2009; Cornell & Khuri, 1996). However multiple issues arise when attempting to implement RSM to solve the Poinsettia problem or a similar problem.

A typical agricultural dose-response problem is depicted in Figure 2. As one factor increases, the response increases, but this increase is limited by a low level of the other factor. In addition, the response is limited; once the optimum level each of the two factors is reached, further increase in the level of either factor does not result in an increased response. RSM methodology attempts to model a surface such as the one depicted in Figure 2 using a 2nd order polynomial. In general, the 2nd order polynomial can be written as:

$$E(y) = \beta_0 + \sum_i \beta_i X_i + \sum_i \beta_{ii} X_i^2 + \sum_{i<j} \beta_{ij} X_i X_j, \text{ where} \tag{1}$$

$E(y)$ denotes the expected value of the response,

β_0 denotes the intercept,

X_i denotes the level of the ith quantitative factor,

β_i denotes the linear regression coefficient for the ith quantitative factor,

β_{ii} denotes the quadratic regression coefficient for the ith quantitative factor,

β_{ij} denotes the linear × linear, or 1st order, regression coefficient for the interaction between the ith and j^{th} quantitative factors

When modeling response patterns such as the one depicted in Figure 2, two issues arise. Often interactions other than two factor 1st order interactions, i.e. linear by quadratic interactions, quadratic by quadratic interactions, or interactions involving more than two factors, exist. The existence of these interactions in plant nutrition research specifically was established by Paparozzi et al. (2005). The second issue is the inadequacy of the 2nd order terms in describing the dose-response pattern. This is evidenced in the single factor case in Figure 3. In

Figure 3, the true optimum dose is over estimated by the quadratic curve. Again, this overestimation is undesirable as it may result economic or environmental loss.

The inadequacy of the 2nd order polynomial is addressed by all the aforementioned texts. While specific implementation of RSM differs from account to account, the basic strategy is the same. First, an exploratory experiment is done using some broad set of combinations of factors levels to find a general region of the factor space in which the optimal combination exists. Further experimentation in this region is then performed to narrow the set possible of factor level combinations even further. This process is repeated until the region of the factor space being studied is small enough that modeling the response as a 2nd order polynomial is an acceptable approximation to the true response.

It is in this process that standard RSM fails researchers when attempting to estimate the optimum factor level combination in situations similar to the Poinsettia problem. The third major characteristic of problems studied by Guo, et al. was the existence of severe restrictions on the amount of time and experimental material available to the researchers. Often, the length of a single run of the experiment is very long, in the Poinsettia case corresponding to the length of the growing season of the plant. Thus, estimating the optimum by repeated experimentation may require years of research. For economic and academic reasons, a study such as this often must be completed in one to two years, restricting the number of experimental runs to one, perhaps two. Therefore alternatives to standard RSM are needed.

The goals of the study by Guo, et al. included examining the performance of alternatives to standard RSM. This included evaluating the performance of models other than the 2nd order polynomial in addition to evaluating the performance of various incomplete factorial designs, deemed necessary due to the restriction of experimental material. This study expands on previous study by including more models and designs in the analysis. In addition, the comparison of models and designs by Guo, et al. utilized simulation based statistics for comparing models. This study aims to demonstrate why simulation based fit statistics are necessary for comparison by also comparing models and designs using traditional fit statistics. All comparison is done via a simulation study. Following the results of the study, a discussion section explains the results and proposes the use of simulation in selecting models and designs when planning an experiment displaying characteristics similar to the Poinsettia problem. In summary, the goals of this study are three fold: first, to evaluate the performance of various designs and models when restrictions are placed on the number of experimental runs and experimental units within a run; second, to propose a method via simulation for the selection of designs and models in the planning of a response surface oriented experiment when standard RSM cannot be applied; and finally, to argue the necessity of simulation based evaluation in the selection of designs and models. The simulation assumes all responses are Gaussian in nature.

2. Alternatives to the 2nd Order Polynomial

The inadequacy of the 2nd order polynomial is not unique to agricultural experiments. Standard RSM address this inadequacy by running a sequence of experiments and progressively restricting the design space to one in which the true response can be reasonably approximated by a 2nd order polynomial. As this is not a reasonable approach to the type of research problem considered here, alternative models to the second order polynomial are needed.

The inadequacy of the 2nd order polynomial results from its inability to effectively describe the diminishing returns effects of factors on the response and the effects of higher order interactions on the response. One approach to modeling diminishing returns, or linear plateau, effects of factors on a response is the use of spline or segmented models (Anderson & Nelson, 1975). However, these models are not conducive to efficient design selection. Alternatively, multifactor nonlinear models can be used. Landes, et al. (2000) presented adaptations of Gompertz and Mitscherlich nonlinear models to fit plant nutrition data described by Macz et al. (1997). In a multifactor setting, the Gompertz and Mitscherlich models, respectively, can be written as:

$$E(y) = \alpha \times \exp\left[\beta \times \exp\left(-\sum_i \gamma_i X_i - \sum_{i<j} \gamma_{ij} X_i X_j - \sum_{i<j<k} \gamma_{ijk} X_i X_j X_k - ... \right) \right] \tag{2}$$

and,

$$E(y) = \alpha - \beta\left(\prod_i \gamma_i^{X_i} \prod_{i<j} \gamma_{ij}^{X_i X_j} \prod_{i<j<k} \gamma_{ijk}^{X_i X_j X_k} ... \right). \tag{3}$$

Here, $E(y)$ and X_i are defined as in the 2nd order polynomial model. In both models, the parameter α denotes the maximum expected response. In the case of the Gompertz model, β denotes the log of proportion of the asymptote that $E(y)$ increases from $X_i=0$ to the values of X_i at which $E(y)$ is effectively at asymptote. While in the case of the Mitscherlich model, β denotes the difference between the maximum expected response and the expected response when all factor levels are set to zero. Finally, in both models, $\gamma_i, \gamma_{ij}, \gamma_{ijk},...$ denote the rates of increase associated with the ith main effects, ij^{th} two-way interactions, ijk^{th} three-way interactions, etc. While not shown explicitly in the model equation, up to n-way interactions can be included for n-factor factorial experiments. Following the presentation of this study by Guo, et al. in 2006, it was suggested a Logistic model may also be a viable option for this modeling scenario. In a multi-factor setting, the logistic model can be parameterized as follows:

$$E(y) = \frac{\alpha}{1 + \exp\left(\beta - \sum_i \gamma_i X_i - \sum_{i<j} \gamma_{ij} X_i X_j - \sum_{i<j<k} \gamma_{ijk} X_i X_j X_k - ... \right)} \tag{4}$$

where all parameter interpretations are similar to those of the Gompertz and Mitscherlich models.

While nonlinear models with asymptotic properties can capture the diminishing returns effects of factors on a quantitative response, the possibility of convergence issues while fitting these models motivates the exploration of linear model alternatives. One such model is the linearized Hoerl model, suggested by Olson, et al. (2001). For a single factor the deterministic

component of the Hoerl model is $E(y) = \exp(\beta_0 + \beta_1 X)\left[X^{\beta_2}\right]$ which can be linearized to $\ln[E(y)] = \beta_0 + \beta_1 X + \beta_2 \ln(X)$. In a multifactor setting, the linearized Hoerl model is:

$$\ln[E(y)] = \beta_0 + \sum_i \left[\beta_{1i} X_i + \beta_{2i} \ln(X_i)\right] + \sum_{i<j}\left[\beta_{1ij} X_i X_j + \beta_{2ij} \ln(X_i)\ln(X_j)\right]$$
$$+ \sum_{i<j<k}\left[\beta_{1ijk} X_i X_j X_k + \beta_{2ijk} \ln(X_i)\ln(X_j)\ln(X_k)\right] + ... \qquad (5)$$

Here, β_0 denotes the intercept, $\beta_{1i}, \beta_{1ij}, \beta_{1ijk}, ...$ denote main effect, 2-way interaction, 3-way interaction, etc. coefficients associated with the literal levels of the factors, and $\beta_{2i}, \beta_{2ij}, \beta_{2ijk}, ...$ denote main effect, 2-way interaction, 3-way interaction, etc. coefficients associated with the log levels of the factors.

The relative fit of the Gompertz model and the Hoerl model in a single factor scenario is displayed in Figure 4. The fitted curves of the Logistic and Mitscherlich models are omitted for clarity because in the single factor case, these two models produce curves similiar to that of the Gompertz model. Both the Gompertz and Hoerl models produce a much better fit than that of the 2nd order polynomial. While the single factor fit of these alternative models is promising, the main objective of this study is to evaluate the performance of the models in a multifactor setting. Figure 5, Figure 6, and Figure 7 depict the 2-factor of the 2nd order polynomial, Hoerl, and Gompertz model respectively, when fit to the surface depicted in Figure 2. As in the single factor case, the Gompertz and Hoerl model provide better approximations of the surface than the 2nd order polynomial.

One major advantage of the alternative models over the 2nd order polynomial is the inclusion of interactions higher than 1st order. It may be argued that augmenting a 2nd order polynomial to include higher order interactions may produce a similar result to that of one of the alternative models discussed thus far. For example, in the three factor case, the addition of a 3-way linear interaction, 2-way quadratic interactions, and a 3-way quadratic interaction may produce much better fit than the usual 2nd order polynomial. Thus, as a final model, the simulation study that follows will also include a model such as this referred to as the "augmented polynomial."

3. Candidate Designs

The alternative models described in the previous section are necessary due to the restriction of the number of experimental runs. Likewise, researchers often faced with severe restrictions on the number of experimental units in any given experimental run. For example, consider a three factor experiment. To estimate the multifactor models described in the previous section, Landes, et al. (2000) suggests a minimum five levels of each factor are required. Thus, a full factorial experiment would require 5^3, or 125, experimental units. But this many experimental units is often the upper limit of a plant science experiment due to limited greenhouse space. In the case that less space is available, or more than three factors are to be tested, a full factorial experiment is not possible. Thus, the choice of a partial factorial design is necessary. As a working example, this paper explores candidate designs in the three factor, five leve, case in

which only 25 experimental units are available. For a more in depth discussion concerning design considerations, including scenarios involving more than three factors and comparisons of designs with respect to optimality criteria, we refer the reader to the original proceedings paper concerning this study (Guo, et al. 2006). The following designs are considered.

All standard response surface designs are based on the 2nd order polynomial. Of these designs, the Central Composite Design (CCD) is the only design that allows for five levels of each factor. It is unclear if the design will be robust to the use of the alternative models described in the previous section. In the 3-factor setting, the design requires only 14 experimental units, thus 11 experimental units are available for replication. The simulation study included three versions of the Central Composite Design: one in which the replicates were placed at the center point of the design, one in which the replicate points were placed at the corner points, and one in which the replicate points were placed at the axial points.

Olson, et al. (2001) proposed an alternative design based on the argument that fractions of 3^N factorial designs are inadequate but can be augmented toallow for estimation of the models described previously. As an example for a three-factor experiment, Olson suggested a possible design based on the "Face Centered Cube" essentially a Central Composite Design with "face" points instead of axial points augmented by a 2^3 factorial. We will henceforth refer to this design as an "augmented face centered cube."

The augmented face centered cube design assumes five levels of each of three factors (0, 1, 2, 3, and 4). Olson, et al. used five levels to guarantee that one would be able to estimate a nonlinear response surface model such as the Gompertz or Mitscherlich. The "face centered cube" uses levels 0, 2, and 4 in a 2^3 factorial using levels 0 and 4 plus six "face points (level 2 of any two factors in conjunction with either level 0 or 4 of the third factor) and a center point (level 2 of all three factors). Embedded within the face centered cube is a 2^3 factorial using levels 1 and 3. This design thus requires a total of 23 treatment combinations out of 125 possible factor level combinations. For the purposes on this study, the two replicate points were placed at the center point of the design.

In addition to the augmented face centered cube, Guo, et al. considered various designs that were either D- or U-optimal with respect to one of the alternative models described in the previous section. Specifically, the original study considered D-optimal designs with respect to the augmented polynomial, Hoerl, and Gompertz models in addition to a U-optimal design with respect to the Mitscherlich model. The inclusion of the Logistic model in the current study motivated the inclusion of D-optimal and U-optimal designs with respect to this model.

Of these "alphabetically optimal" designs, only the D-optimal designs with respect to the augmented polynomial and Hoerl models are exact designs. The remaining designs are approximately D-optimal and U-optimal. This is due to the dependency of the remaining designs on the hypothesized values of the parameters of their respective nonlinear models. Thus the designs are only locally optimal with respect to the parameter values used in their construction. Specific details concerning the construction of these designs can be found in the original proceedings paper or in the text by Atkinson, Donev, and Tobias (2007). The parameter values used for construction of the designs were obtained by fitting each model to the surface to be used in the simulation study, depicted in Figure 10. The resulting designs can be found in Table 1.

The designs in this section can be placed into two broad categories: those requiring prior information about the surface to be estimated and those requiring no prior information. The

approximate D-optimal designs with respect to the Gompertz and Logistic models and the U-optimal designs with respect to the Logistic and Mitscherlich models require information concerning the surface via the dependency of the designs on hypothesized values for the model parameters. The Central Composite Designs, augmented face centered cube design, and D-optimal designs with respect to the augmented polynomial and Hoerl models require no prior information concerning the surface to be modeled. Therefore, it is of particular interest if the use of the designs based on hypothesized parameter values increases the accuracy of the results of a study such as the one described by the simulation study.

4. Simulation Study and Results

A simulation study was conducted to explore two aspects of the behavior of the models described in section two under the designs described in section three: first, how well a each model/design combination reproduces a surface similar to that in Figure 2, and second, the convergence rate of the various nonlinear models under each candidate design. An "ideal" population, or true surface, based on the situation portrayed in Figure 2 was constructed. This expected response of this surface is a result of the levels of three factors: X_1, X_2, and X_3, each with levels 0, 1, 2, 3, and 4, where level 0 for each factor was defined to be "limiting" and level 4 was defined to be the maximum. Figure 11 depicts the true surface.

Under each of the ten designs described in the previous section, 1500 simulated data sets were generated. Simulated observations were generated by taking the response value for the true for each given $X_1 \times X_2 \times X_3$ treatment combination and adding a deviate generated from a normal distribution with a mean of 0 and a variance of 3. The responses in the ideal population range from 5 for the limiting treatment levels to a maximum of 14, thus the variance corresponds roughly to a 10-20% coefficient of variation, typical for plant science studies.

Each model was estimated using each simulated data set. The 2nd order polynomial, augmented polynomial, and Hoerl models were estimated using PROC GLM (SAS, 2008). For the Hoerl model, the right-hand side of equation (4) was fit to directly to the response variable, Y, in lieu of fitting the model to log(Y). This was done based on the researchers' experience that applying the right-hand side of the linearized Hoerl directly to Y produces better results (Guo, et al. 2006). The Gompertz, Mitscherlich and Logistic models were fit using PROC NLIN (SAS, 2008).

The models were evaluated using two classes of statistics. Two traditional fit statistics, Aikaike's Information Criteria (corrected) (AICC) and sums of squares for error (SSE), were included to describe the performance of the models as observed by a researcher attempting to determine the ideal model for predicting the surface. The median AICC and SSE for each model/design combination can be found in Table 2. Two simulation based fit statistics were also included, both calculated to compare the predicted surface of each model to the true surface. For each model fit to each data set, the estimated coefficients were used to compute predicted responses for all 125 $X_1 \times X_2 \times X_3$ treatment combinations in the population. The two simulation based statistics calculated were the correlation between the true and predicted responses and the sum of squared prediction errors (SSPE) over the 125 $X_1 \times X_2 \times X_3$ treatment combinations. Because PROC NLIN is an iterative procedure, it failed to converge for some of the simulated

data sets when fitting the nonlinear models. As some designs were more prone to non-convergence than others, the convergence percentage across the 1500 experiments was an additional criterion for the simulation study. The median correlation, median SSPE, and convergence percentages for each model and design combination can be found in Table 3.

Table 2 shows the AICC and SSE results. Other than the 2nd order polynomial model yielding the highest AICC and SSE under all designs, no clear patterns for the alternative models emerge. Across all designs, the Hoerl model provides the smallest SSE. However, the Hoerl model requires more parameters than the nonlinear models. Thus, the nonlinear models provide the smallest AICC. Using AICC and SSE as fit criteria, the individual performances of the nonlinear models are virtually indistinguishable. However, as will become clear in the next paragraph, AICC and SSE present a misleading picture of model/design performance.

Consider the results of the simulation study (Table 2) under the D-optimal Hoerl design using the Hoerl model and the Logistic model. Sums of squares for error, 29.33 and 50.96 for the Hoerl model and Logistic model respectively, indicate the two models models fit the data at considerably different levels of adequacy. Under ideal conditions, this would indicate that the Hoerl model predicts the true surface more accurately than the Logistic model. However, the predicted surfaces for the two models (Figure 12), averaged over the 1500 simulated experiments, appear to predict the true surface almost equally well, with the Logistic model perhaps predicting somewhat more accurately. While this may seem a minor issue, this discrepancy between fit statistics and the accuracy of surface prediction can take extreme forms. Consider the results under the Central Composite Design with replication at the center point. Sums of squares for error, 27.94 and 47.49 for the Hoerl model and Logistic model respectively, indicate the two models perform approximately the same under this design as under the D-optimal Hoerl design. However, average predicted surfaces under this design (Figure 13) indicate this is not true. Both models produce predicted surfaces that differ considerably from the true surface. In fact, the surface produced by the Hoerl model under the Central Composite Design does not produce a surface that even resembles the true surface. While this poor prediction is not apparent when traditional fit statistics are used, it is apparent when simulation based statistics such as SSPE are utilized. For example, the SSPE values under the D-optimal Hoerl design for the Hoerl and Logistic models respectively were approximately 197 and 148. Under the Central Composite Design, the poor prediction is reflected by SSPE values of approximately 220,000,000 for the Hoerl model and 1260 for the Logistic model. These results indicate model/design performance is better described by simulation based fit statistics.

The results as summarized by SSPE, correlation, and convergence percentages are described in Table 3. An analysis of these results yields several conclusions. The performance of the Central Composite Designs is unacceptable. The median SSPE values across all models are orders of magnitude larger than those under other designs. The D-optimal design with respect to the Hoerl appears to be the most robust design. The three nonlinear models and the linearized Hoerl model all produce results with SSPE less than 200 and correlation between the true response and predicted response greater than 0.95. The convergence percentage for each nonlinear model is at or close to 100%. The three nonlinear models perform nearly as well with the augmented face centered cube as with the D-optimal Hoerl design. The augmented polynomial and the 2nd order polynomial are progressively worse. Under the D-optimal design with respect to the augmented polynomial, the augmented polynomial model's performance is

indeed optimized, but the results are noticeably worse that that the Hoerl and Gompertz. In addition, although the Gompertz model results in the best fit when it converges, the convergence rate is only 85.3%.

The design/model combinations with similar SSPE values can be further analyzed using levelplots, which can be produced using software such as the Lattice package in the software program R. This type of plot, in conjunction with squared prediction error, can provide information concerning how a model or set of models performs in certain regions of the design space under a hypothesized surface. An example is given in Figure 11.

In lieu of summing the squared prediction errors to form sum of squared prediction error, squared prediction error (SPE) can be analyzed for each of the 125 factor level combinations over the 1500 experiments under each design and model combination. In the simulation study, the D-optimal Hoerl design appeared to be robust to model selection. This levelplot provides information concerning the performance of the Logistic, Hoerl, and Gompertz models under this design. The results for each model are arranged by row. The levels of X_1 and X_2 increase from left to right, while the level of X_3 increases from the bottom of the plot to the top. The value of interest for each combination of X_1, X_2, and X_3 is the median SPE over the 1500 experiments simulated under the D-optimal Hoerl design. Small values of SPE, displayed by a light color with respect to the spectrum displayed at the right edge of the plot, indicated good fit for that combination of factor levels. Likewise, dark color with respect to the spectrum, indicates poor fit for that combination of factor levels. The actual design points for the D-optimal Hoerl design are overlaid on the plot to display in the relationship between design point selection and model performance of each combination of factor levels.

In this particular example, the large regions of light color for the Logistic and Gompertz models indicate the models tend to fit the surface well for the majority of the factor level combinations. However, the particularly dark regions indicate the models tend to fail to adequately describe the hypothesized surface for these factor level combinations. In contrast, the Hoerl model does not provide the best fit for the majority of factor level combinations, but does tends to predict all factor level combinations equally well. In the context of researcher attempting to select a design and a set of candidate models for a nonlinear dose-response experiment, the levelplot provides the researcher with additional information concerning the potential performance of various candidate models that may be used to analyze data collected from such an experiment.

As a final point, the designs ostensibly optimized for the nonlinear models in practice perform only marginally better than the D-optimal Hoerl design, and only then for their intended models. The lack of results under these approximately optimal designs for nonlinear models is amplified by the fact that the true parameter values, which would not be available in practice, were used to produce the designs. Thus nothing is gained by attempting to tailor a design to a particular nonlinear model. These results and their limitations are explained further in the discussion section.

4. Discussion

The goals of this expansion of the study by Guo, et al. were three fold: first, to evaluate the performance of various designs and models when restrictions are placed on the time and

experimental material available; second, to propose a method via simulation for the selection of designs and models in the planning of a response surface oriented experiment when standard RSM cannot be applied; and finally, to argue the necessity of simulation based evaluation in the selection of designs and models.

The results of the simulation study suggest that the nonlinear Gompertz, Logistic, and Mitscherlich models and the linearized form of the Hoerl model used in conjunction with a D-optimal response surface design optimized for the Hoerl provide a good general purpose approach to exploratory investigation of nonlinear, diminishing return response surfaces in agriculture when resources are limited. While the simulation study is limited to a specific scenario, it is a typical scenario for many kinds of agricultural research from a variety of disciplines. In any scenario where standard RSM does not appear to be applicable due to time and experimental material restrictions, the choice of an appropriate experimental design and useful models is a difficult one. Thus, we propose the simulation study as the first step in the design of an experiment similar to the one described.

The true surface of interest is unknown to the researchers prior to the experiment. But agricultural researchers can often provide sufficient information concerning the relationship between the factors of interest and the response variable to the consulting statistician. Thus, the process of exploring designs and models via simulation should start with a "sketching out" of a hypothesized surface based on information provided by the agricultural researcher. The shape of the hypothesized surface can be used determine candidate models and designs for the study. Once these models and designs have been identified, data sets can be simulated under each of the candidate designs using the hypothesized surface as the true response. Finally, in a manner similar to the previously described simulation study, the designs and models can be analyzed by fitting the candidate models to each of the simulated data sets and producing simulation based fit statistics such as SSPE, correlation, and convergence percentage.

In the previously described simulation study, no one model appeared to outperform any other across all, or even many, of the possible designs. And in the framework of a simulation study for the purpose of planning an experiment, it would be premature to select only one model as a candidate for estimating the true surface. However, as in the case of the D-optimal Hoerl design used in the simulation study, one design may stand out as fairly robust to model selection. That is, the design produces acceptable model fit across a range of possible candidate models. If so, this can lead the researcher to an appropriate design for use in the experiment. Or, at a minimum, this process may eliminate designs that perform poorly across many models under consideration.

The evaluation of models and designs under a hypothesized surface need not be restricted to statistics describing the overall fit of a model under a design. Often, multiple research questions concerning the surface of interest need to be answer in the course of a single study. As a result, the estimation of specific regions of the surface may be of more importance than others. For example, the researchers may be interested in the response in the region of initial change with respect to the increasing values of the factors, in the region of some particular combination of factor levels, or in the region where the response begins to plateau. While the exact region of interest may be unknown, it too can be hypothesized providing the consulting statistician with more information to include in the construction of the design. Using the method illustrated in the

results section, this information can be utilized in conjunction with levelplots to determine the accuracy of various models in predicting a certain region of the surface of interest.

The original study, and this expansion of the study, utilized simulation based statistics to evaluate the performance of various models and designs. The third goal of this study is to motivate the use of these simulation based statistics in lieu of traditional fit statistics by demonstrating the inadequacy of traditional fit statistics when utilizing simulation to design a dose response experiment. As the examples in the results section illustrated, traditional fit statistics such as SSE and AICC can provide misleading information when determining useful models and designs. Hence, simulation based statistics such as SSPE, correlation, and convergence percentage are necessary to provide an accurate description of model performance.

The final conclusions of this study are two-fold. First, in experiments involving multiple factors with diminishing returns or dose-response relationships with the response, traditional response surface methodology is not applicable and alternative methods are needed. Traditional designs and models, such as the Central Composite Design and the 2nd order polynomial, cannot adequately describe the response when time and experimental material are limited. The results of this simulation study indicate a design based on the linearized Hoerl model may be useful in this context and the use of the Hoerl model or a nonlinear model may applicable. Second, simulating the experiment being planned under a hypothesized surface constructed using information provided the agricultural scientists can be a very useful process in determining what designs and models may be applicable to the situation at hand. Simple statistics and plots, such as correlation, squared prediction error, and the levelplot, can be used to inform the researchers of the performance of models and designs while planning an experiment under time and material restrictions.

5. Summary

This research expanded on the work by Guo, et al. that investigated design and analysis of nonlinear dose response experiments. Specifically these experiments must be designed under severe constraints of time and experimental material. In addition, the response is measure with respect to several experimental factors, the effect of each being limiting in nature. Traditional response surface methodology is inadequate when designing an experiment under these conditions. Guo et al. concluded a design based on a candidate set of 5 levels per factor, constructed to be D-optimum with respect to a linearized Hoerl model, is efficient when model estimates are produced using a Gompertz model. The expansion of this study to include both Central Composite Designs and a Logistic model resulted in similar findings. While the results of one simulation study cannot be universally accepted for all dose-response scenarios, a reasonably efficient design for any dose-response scenario can be produced using a simulation study, assuming some information concerning the surface of interest is available.

6. References

Anderson, R.L. and L. A. Nelson. 1975. A family of models involving intersecting straight lines and concomitant experimental designs used in evaluating response to fertilizer nutrients. *Biometrics.* **31:** 303-318.

Atkinson, A. C., A. N. Donev, and R.D. Tobias, 2007, *Optimum Experimental Designs, with SAS,* Oxford University Press.

Box, G.E.P., W.G. Hunter, and J.S. Hunter, 2005. *Statistics for Experimenters: Design Innovation, and Discovery, 2nd Ed.*, New York: John Wiley & Sons.

Cornell, J.A. and A.I. Khuri. 1996. *Response surfaces: Designs and analyses, 2nd Ed.* New York: Marcel-Dekker.

Guo, S., W.W. Stroup, E.T. Paparozzi and M.E. Conley. 2006. A comparison of models and designs for experiments with nonlinear dose-response relationships. *In: Proceedings 18th Conference on Applied Statistics in Agriculture.* Department of Statistics. Kansas State University. Manhattan, KS.

McCullagh, P., Regression models for ordinal data, *Journal of the Royal Statistical Society, Series B*, **42:** 109-142.

Myers, R.C. and D.M. Montgomery. 2009. *Response surface methodology: process and product optimization using designed experiments, 3rd Ed.* New York: Wiley-Interscience.

Olson, L. M., W. W. Stroup, E. T. Paparozzi and M. E. Conley. 2001. Model building in multi-factor plant nutrition experiments. pp. 183-206. *In: Proceedings 12th Conference on Applied Statistics in Agriculture.* Department of Statistics. Kansas State University. Manhattan, KS.

Paparozzi, E.T., W.W. Stroup, and M.E. Conley. 2005. How to investigate four-way nutrient interactions in plants: a new look at response surface methods. *Journal of the American Society for Horticultural Science,* **130**: 459-468.

SAS Institute Inc. 2008. *SAS/STAT user's guide, Vesion 9.2*. Cary, NC: SAS Institute Inc.

Figure 1. Visual Response of Poinsettia to Levels of N and S Applied

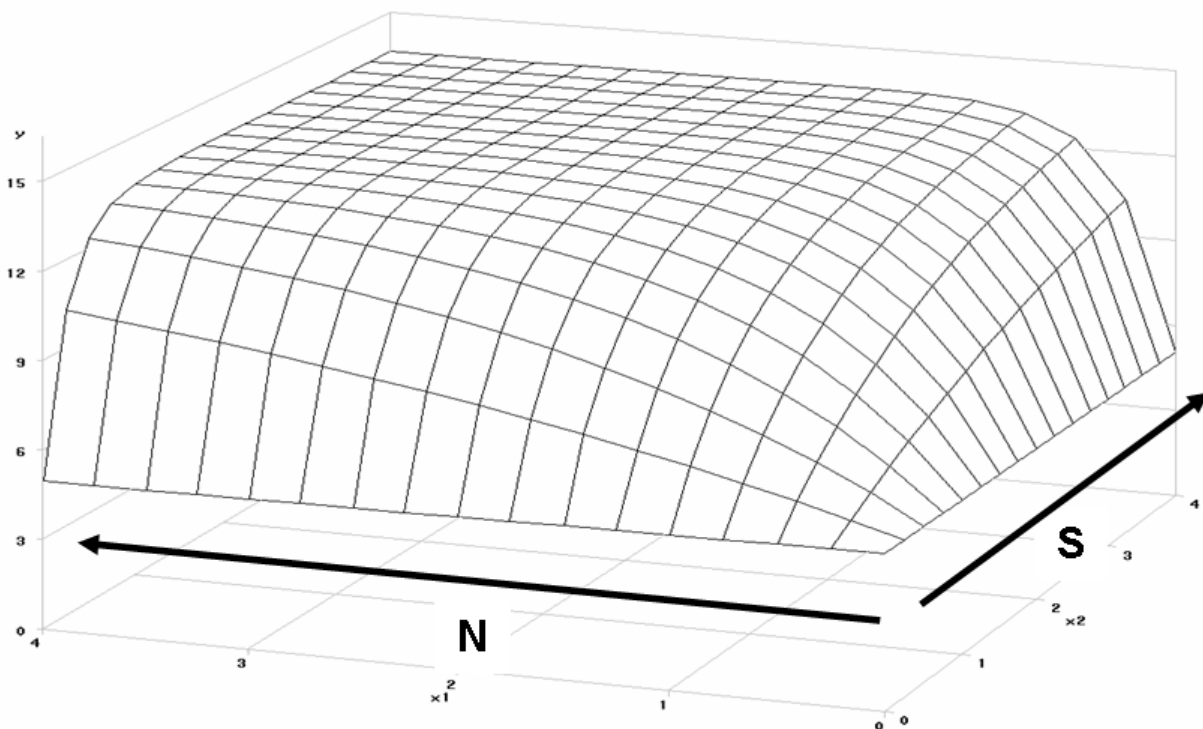**Figure 2.** Plot of Poinsettia Dose-Response Profile

Figure 3. Quadratic Approximation of Diminishing Returns Response
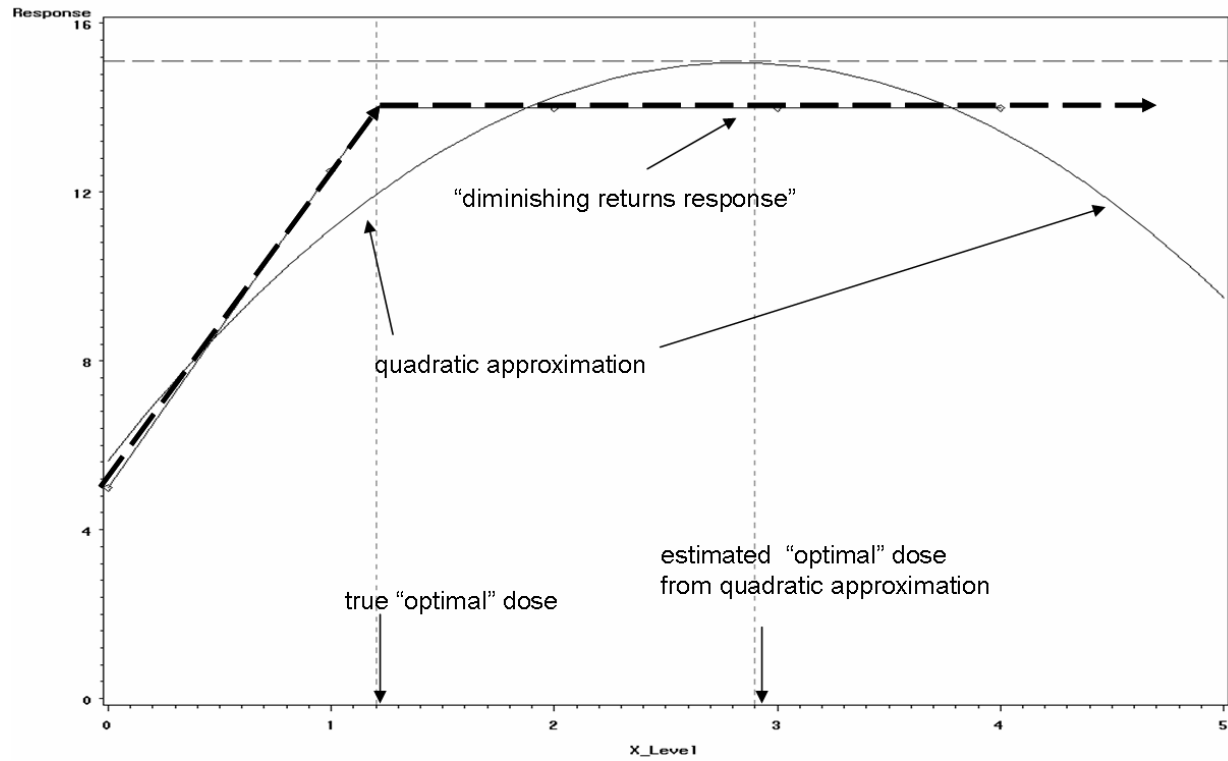

Figure 4. Gompertz and Hoerl Approximations of Diminishing Returns Response
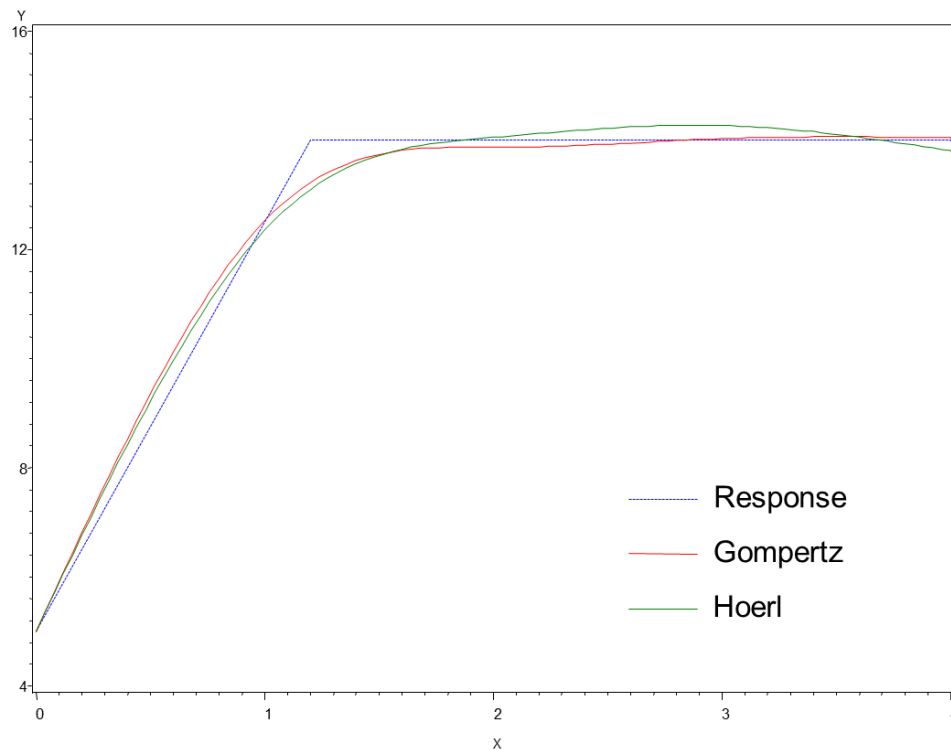
Figure 5. Two-factor 2nd Order Polynomial fit of Figure 2 Dose-Response Surface
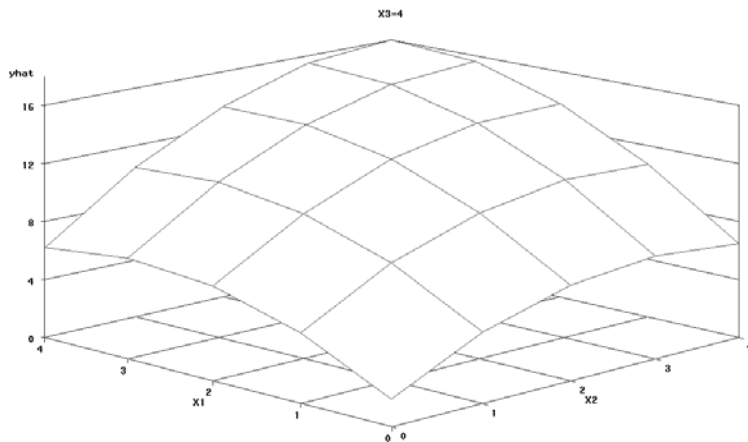
Figure 6. Two-factor Hoerl fit of Figure 2 Dose-Response Surface
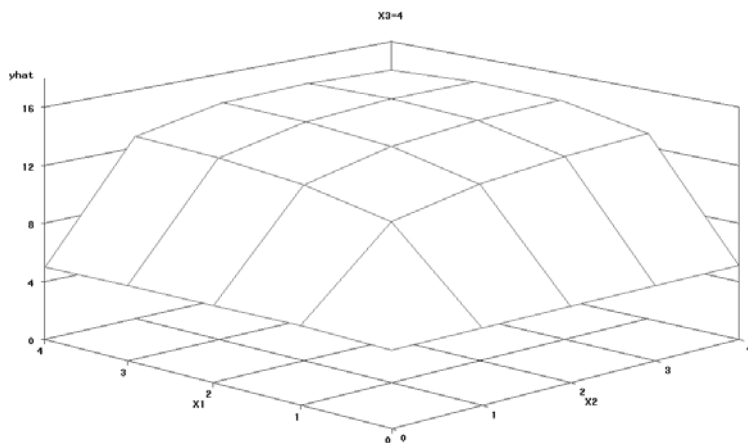
Figure 7. Two-factor Gompertz fit of Figure 2 Dose-Response Surface
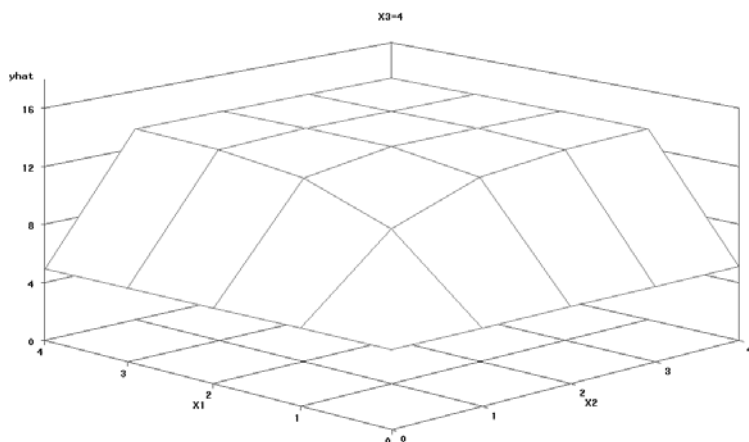
Figure 8. Two-factor augmented Polynomial fit of Figure 2 Dose-Response Surface
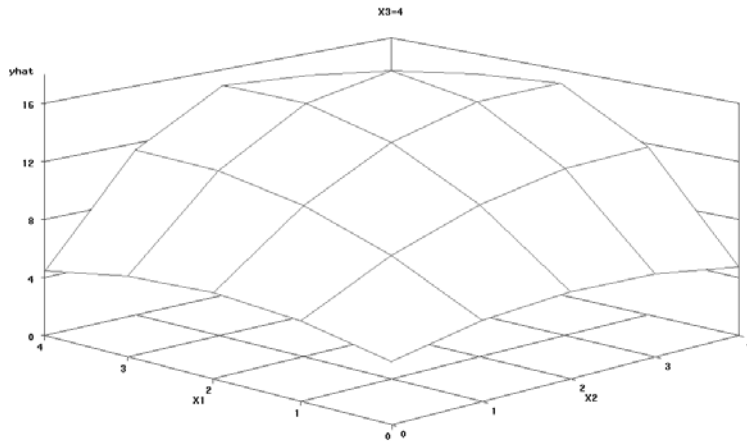
Figure 9. Olson's "CELEPSO" Three-Factor Response Surface Design

Figure 10. "Ideal" Population for Simulation Study – Three Factor "Diminishing Returns" Nonlinear Response Surface

Figure 11. Squared Prediciton Error "Levelplot" under the D-optimal Hoerl Design

Figure 12. Predicted Surfaces under the D-optimal Hoerl Design

Figure 13. Predicted Surfaces under the Central Composite Design (replication at center point)

Table 1. Designs generated as optimal with respect to models under study.

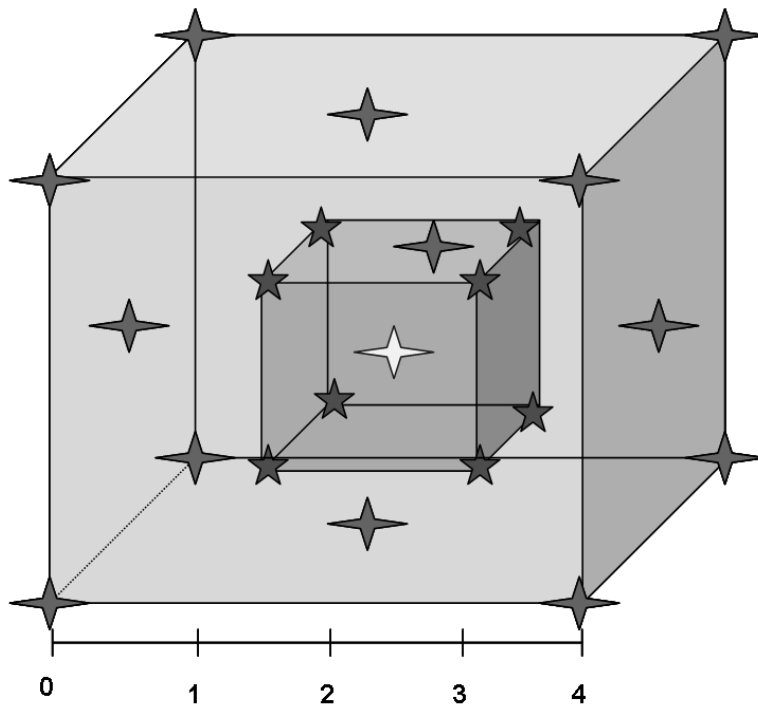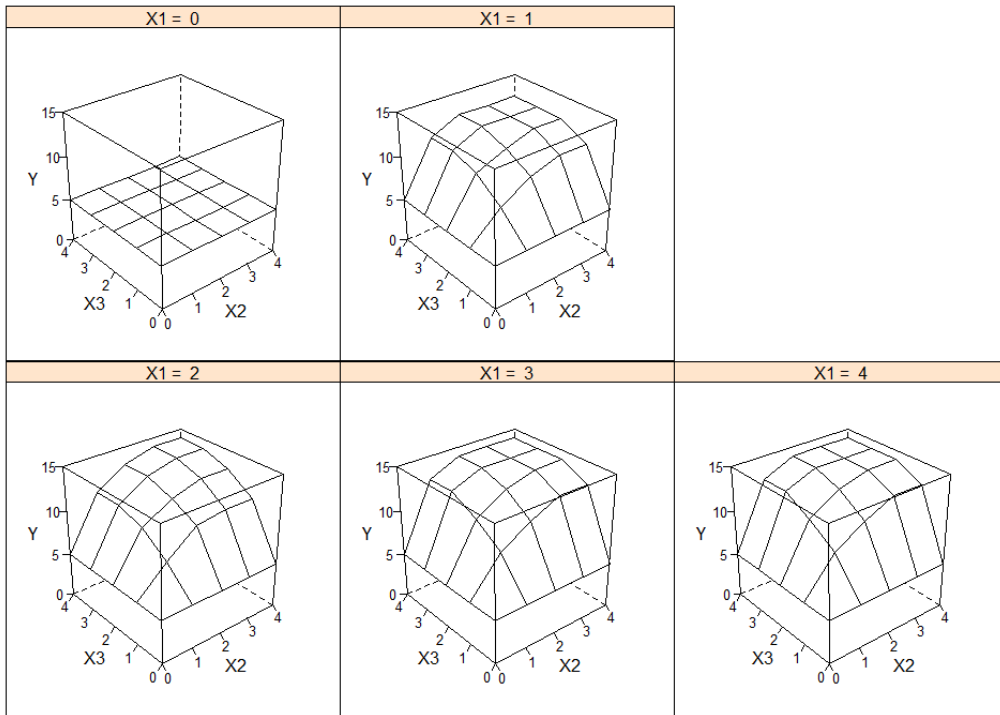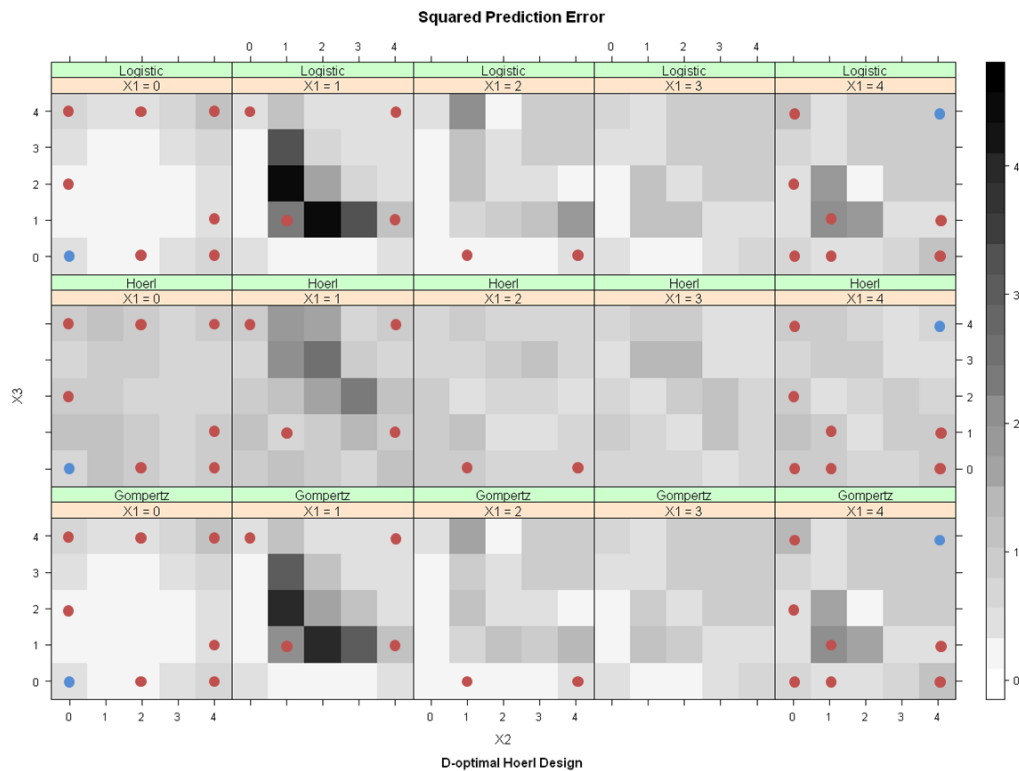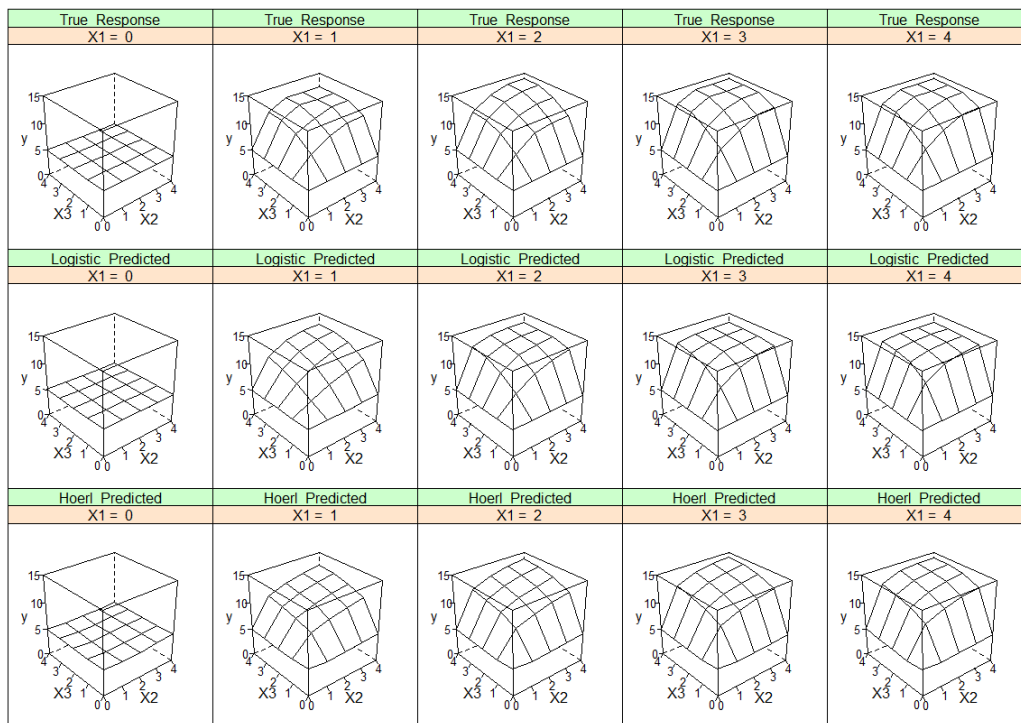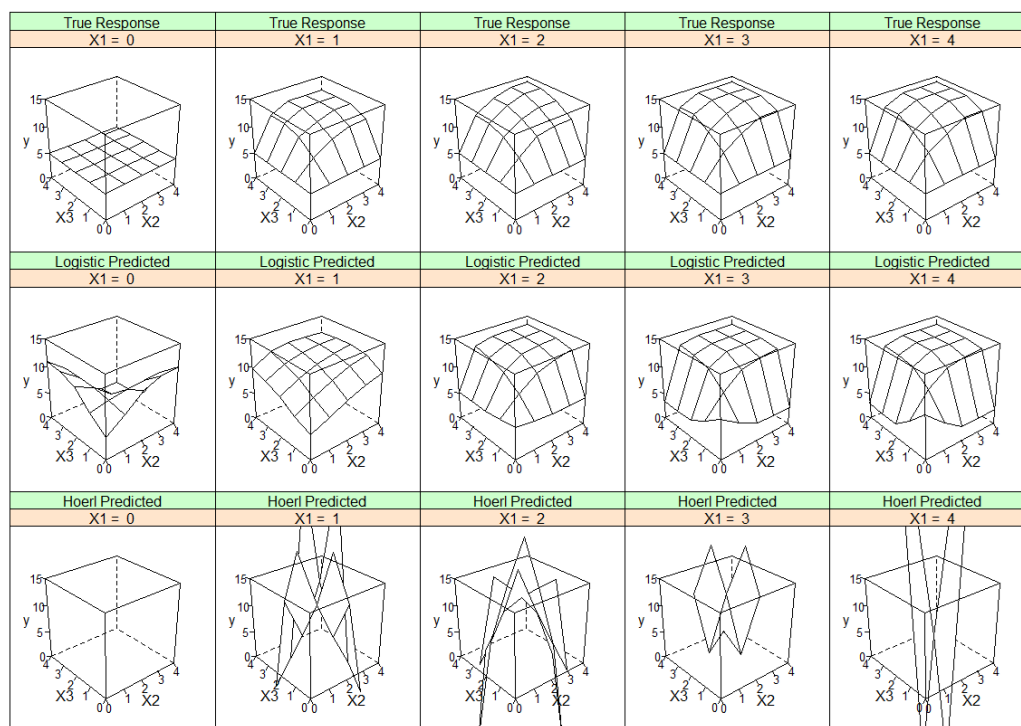| D-optimal 2nd Order Poly.			D-optimal Hoerl			D-optimal Gompertz			D-optimal Logistic			U-optimal Logistic			U-optimal Mitscherlich		
X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	4
0	0	4	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
0	2	2	0	0	2	0	0	0	0	0	0	0	2	2	0	2	2
0	2	4	0	0	4	0	0	4	0	0	4	0	3	3	0	3	3
0	4	0	0	2	0	0	0	4	0	0	4	0	3	4	0	3	4
0	4	2	0	2	4	0	4	0	0	0	4	0	4	0	0	4	0
0	4	4	0	4	0	0	4	0	0	4	0	0	4	4	0	4	4
2	0	0	0	4	1	0	4	0	0	4	0	1	1	3	1	0	2
2	0	2	0	4	4	0	4	4	0	4	4	1	3	1	1	1	3
2	0	4	1	0	4	0	4	4	0	4	4	1	3	2	1	2	0
2	2	0	1	1	1	0	4	4	0	4	4	2	0	2	1	3	1
2	2	2	1	4	1	1	1	3	1	1	3	2	1	0	2	0	2
2	2	2	1	4	4	1	3	1	1	3	1	2	2	0	2	2	0
2	2	4	2	0	0	3	1	1	2	4	2	2	2	2	2	2	2
2	4	2	2	4	0	3	3	3	3	3	3	3	0	3	3	0	2
2	4	4	4	0	0	4	0	0	4	0	0	3	0	4	3	0	3
4	0	0	4	0	2	4	0	0	4	0	0	3	1	1	3	0	4
4	0	2	4	0	4	4	0	4	4	0	4	3	1	2	3	1	1
4	0	4	4	1	0	4	0	4	4	0	4	3	2	2	3	3	0
4	2	0	4	1	1	4	0	4	4	0	4	3	3	0	4	0	0
4	2	2	4	1	4	4	4	0	4	1	1	3	3	3	4	0	4
4	2	4	4	4	0	4	4	0	4	4	0	3	4	0	4	2	2
4	4	0	4	4	1	4	4	0	4	4	0	4	0	0	4	3	0
4	4	2	4	4	4	4	4	4	4	4	0	4	0	4	4	4	0
4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4

Table 2. Small sample performance using traditional fit statistics

Design	Criteria	Linear Models			Nonlinear Models		
		2nd Order Polynomial	Augmented Polynomial	Hoerl	Gompertz	Logistic	Mitscherlich
Augmented FBC	SSE	108.78	50.40	29.62	52.14	51.86	50.44
	AICC	79.07	138.95	104.21	54.09	53.96	53.26
Central Composite Design 1	SSE	72.74	27.94	27.94	46.32	47.49	42.84
	AICC	69.01	124.19	102.77	51.13	51.75	49.17
Central Composite Design 2	SSE	65.66	27.95	27.95	48.47	49.95	43.96
	AICC	66.44	124.21	102.78	52.26	53.01	49.81
Central Composite Design 3	SSE	72.34	27.94	27.89	48.26	48.70	52.38
	AICC	68.87	124.21	102.73	52.15	43.80	49.71
D-optimal Gompertz	SSE	75.15	36.49	36.45	50.59	48.19	46.35
	AICC	69.82	130.88	109.42	53.34	52.12	51.15
D-optimal Hoerl	SSE	101.00	38.10	29.33	50.10	50.96	49.64
	AICC	77.21	131.96	103.99	53.09	53.51	52.86
D-optimal Logistic	SSE	64.24	34.23	34.16	53.33	51.99	56.45
	AICC	65.90	129.28	107.81	54.65	54.02	56.07
D-optimal 2nd Order Polynomial	SSE	66.18	27.84	27.84	45.07	45.69	44.93
	AICC	66.64	124.12	102.69	50.44	50.78	50.37
U-optimal Logistic	SSE	128.54	42.14	29.08	51.29	51.25	50.84
	AICC	83.24	134.47	103.77	53.68	53.66	53.46
U-optimal Mitscherlich	SSE	101.07	37.17	28.92	49.98	50.68	49.57
	AICC	77.23	131.35	103.65	53.03	53.38	52.83

Table 3. Small sample performance using simulation based fit statistics

Design	Criteria	Linear Models			Nonlinear Models		
		2nd Order Polynomial	Augmented Polynomial	Hoerl	Gompertz	Logistic	Mitscherlich
Augmented FBC	SSPE	809.71	809.71	427.72	165.92	163.95	146.54
	Corr	0.73	0.73	0.88	0.95	0.95	0.96
	Conv. %				100.0%	100.0%	100.0%
Central Composite Design 1	SSPE	1259.28	25179826.14	221399102.77	1270.67	1260.26	614539.78
	Corr	0.70	-0.09	0.01	0.74	0.73	0.02
	Conv. %				69.4%	97.0%	12.8%
Central Composite Design 2	SSPE	1548.58	26697720.93	249107665.58	1263.78	1263.65	1701690.37
	Corr	0.68	-0.08	0.02	0.73	0.71	0.02
	Conv. %				64.1%	95.0%	13.6%
Central Composite Design 3	SSPE	1622.41	27250732.08	277058178.98	1348.23	1281.09	1070243.38
	Corr	0.72	-0.09	0.02	0.72	0.72	0.02
	Conv. %				73.5%	96.0%	13.6%
D-optimal Gompertz	SSPE	1428.33	1735.15	3613.00	145.50	152.88	1240.67
	Corr	0.58	0.52	0.52	0.95	0.96	0.79
	Conv. %				99.9%	99.7%	78.1%
D-optimal Hoerl	SSPE	545.87	492.62	196.83	146.52	148.07	142.22
	Corr	0.83	0.89	0.94	0.95	0.95	0.96
	Conv. %				100.0%	100.0%	100.0%
D-optimal Logistic	SSPE	1276.96	1891.57	3919.20	127.59	134.37	934.52
	Corr	0.59	0.47	0.47	0.96	0.96	0.80
	Conv. %				96.6%	93.7%	75.7%
D-optimal 2nd Order Polynomial	SSPE	478.15	382.31	300.63	162.40	179.43	344.69
	Corr	0.86	0.89	0.90	0.95	0.95	0.90
	Conv. %				100.0%	98.4%	71.9%
U-optimal Mitscherlich	SSPE	767.32	1243.03	3013.63	165.22	171.06	160.87
	Corr	0.81	0.86	0.46	0.95	0.95	0.95
	Conv. %				99.1%	97.8%	97.7%
U-optimal Logistic	SSPE	842.24	7091.75	3053.06	168.12	173.53	169.63
	Corr	0.75	0.27	0.48	0.95	0.95	0.34
	Conv. %				98.7%	96.5%	94.9%