An Overview of ADX Menu System and Design of Experiments

Almira A. Dumaual & Timothy E. Cooley, Boehringer Mannheim Corporation

ABSTRACT

The ADX menu system in SAS™ aids in the design and analysis of experiments. However, a complete understanding of this software cannot be accomplished without some background information on design of experiments. This paper will provide both an overview of design of experiments and ADX through the use of an example and will emphasize how the ADX menu system can simplify the process from the development of an experiment to the analysis of its data.

INTRODUCTION

Design of experiments (DOE) is a strategy that assists researchers in the identification and characterization of the important factors in a process. DOE can improve research and manufacturing productivity by reducing the cost, time to market, and scrap of a product. It also allows a better understanding of the process in question and makes the product robust.

The SAS™ ADX Menu System is a statistical tool that assists both the novice and expert in this process by providing prompts and menus that act as a guide in the selection and analysis of an experimental design.

This paper will introduce the four phases of DOE and then give an example of a response surface design created and analyzed in ADX.

DESIGN OF EXPERIMENTS

Background.

GENERAL.
The four phases of the DOE strategy can be characterized by the experiments implemented within them. These phases involve the design and analysis of four types of experiments: screening, interaction, response surface, and confirmation experiments. In contrast to changing one variable at a time, DOE starts with the investigation of many variables at once and then proceeds to characterize the process in question by removing the insignificant factors from the model and optimizing its mean response.

The factors are identified as those variables which are believed to have an effect on the response and their levels are predetermined from historical data. The response should be quantitative and relate directly and clearly to the process in question.

A mathematical model is identified for each phase of DOE (a simple polynomial, usually second order or less) and solved using the least squares regression method.

CONFOUNDING PATTERNS AND RESOLUTIONS.
All possible interactions among the factors can be estimated in a full factorial design. However, in other designs, the ability to estimate some of the higher order interactions are lost in order to reduce the sample size of the experiment. These interactions are *confounded* with other terms in the model, thus their independent estimates cannot be obtained. For example, if the factor A is confounded with the interaction BC, then it cannot be determined whether the effect of A calculated in the analysis is due to A or BC or some combination of the two terms.

The classes of confounding are described by *resolutions*:

- Resolution III - main effects are confounded with two-way interactions but are not confounded with each other.
- Resolution IV - main effects are not confounded with each other or with two-way interactions, but two-way interactions are confounded with each other.
- Resolution V - no main effects or two-way interactions are confounded with each other, assuming higher order interactions are absent.

Typically, screening designs have Resolution III, interaction designs have Resolution IV, response surface designs have Resolution V.

REPLICATION AND CENTER POINTS.
Replication of observations and center points in a design is advantageous for statistical analysis. Replication provides additional degrees of freedom for the estimation of error variance which results in greater power when testing the significance of effects. It also allows for an independent estimate of the experimental error. Center points provide information about the interior of the design and check for curvature.

Some designs generated in ADX do not include replications. Therefore, replications may need to be added as observations in the experiment.

The Screening Phase.
The first phase of the design process is the screening phase. The purpose of screening experiments is to select, from a candidate set of *m* variables, those parameters that have a significant affect on the mean response. They also can determine how changes in the levels of the important factors affect the performance. The regression model for screening designs has the general form:

$$y=\mu+\sum_{i=1}^{m}\beta_i X_i+\epsilon$$

where only the main effects β_i are estimable. A few of the experimental designs typically used for screening are:

- Fractional factorials
- Plackett-Burman designs
- Taguchi designs

These designs are two-level designs with high and low factor levels based on past information, commonly set at -1 and +1, after perhaps a suitable change in location and scale of measurement.

These designs allow for the analysis of many factors using very few observations so that the number of important variables can be reduced and investigated more thoroughly in subsequent experiments.

However, in this DOE phase, two-way (or higher) interactions cannot be estimated and it is most likely that optimization of the process has not been reached. Therefore, the second phase should be implemented.

The Interaction Phase.
Interaction designs are used for problem reduction in experiments and have a similar purpose as screening designs. The model is different from that of screening designs, however, because they evaluate two-way (or higher order) interactions which are believed to be important from past experience. The regression model that is defined for these designs has the general form:

$$y=\mu+\sum_{i=1}^{m}\beta_i X_i+\sum_{i\neq j}^{m}\sum\beta_{ij}X_iX_j+\epsilon$$

All main effects and some two-way interactions are estimable, while other two-way interactions may be confounded with each other.

A list of several interaction designs follows:

- Full factorials
- Fractional factorials
- Taguchi designs

Similar to screening designs, interaction designs are also two-level designs.

Although some interactions are accounted for in this mathematical model, this phase may still not fully characterize the process in question. Thus, the third DOE phase should be executed.

The Optimization Phase.
Response surface designs are otherwise known as optimization experiments. As suggested, the purpose of these designs is to build a predictive model for the process being observed by defining the specific optimal values for the experimental factors in question. The mathematical models used to define these designs are usually second order polynomial models. They have the general form:

$$y=\mu+\sum_{i=1}^{m}\beta_i X_i+\sum_{i\neq j}^{m}\sum\beta_{ij}X_iX_j+\sum_{k=1}^{m}\beta_{kk}X_k^2+\epsilon$$

In this model the specified main effects, interactions, and quadratic terms are estimable.

Response surface designs are three-level designs with center points. Some of these designs are:

- Box-Behnken designs
- Central Composite designs
- Central Composite face designs

Although these designs optimize the mean response of the process, the DOE strategy requires that a further and final experiment is implemented.

The Confirmation Phase.

The confirmation process involves testing the optimal model. First, the response of the model is calculated from the prediction equation using specific factor levels. Then, an actual experiment is run using the same settings. The predictive model is confirmed if the response from the experiment is comparable to that calculated from the equation.

ADX EXAMPLE.

The following example shows how the ADX Menu System in SAS™ creates and analyzes a response surface design.

Background of experiment.
This experiment has three variables: buffer, pH, and ionic strength. The regression model that defines the design is:

$$\begin{aligned} y = \mu &+ \beta_1 buffer + \beta_2 pH + \beta_3 ion \\ &+ \beta_{12} buffer*pH + \beta_{13} buffer*ion \\ &+ \beta_{23} pH*ion + \beta_{11}^2 buffer^2 + \beta_{22}^2 pH^2 \\ &+ \beta_{33}^2 ion^2 + \epsilon \end{aligned}$$

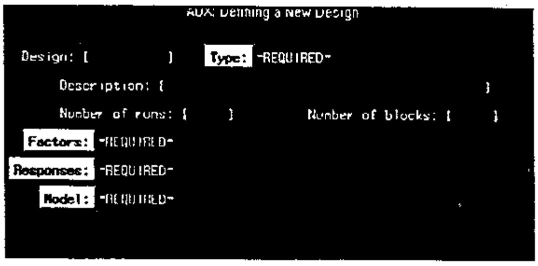
The objective of this experiment is to optimize, in both directions, the response y.

ADX Menu System.

DESIGN CREATION.
The ADX Menu System is invoked by selecting PLANNING TOOLS in SAS/ASSIST and then selecting DESIGN OF EXP in the Planning Tools Menu.

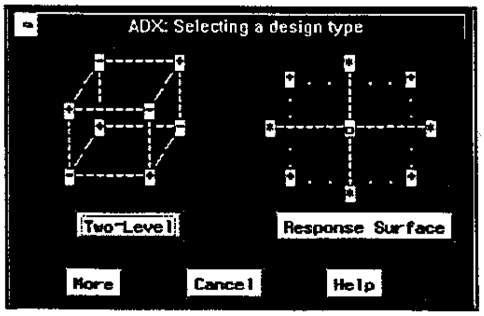
The *design directory* window appears:

In order to create a design in ADX, select New Design under the File menu option in the *design directory* window, and then select Define and create from the New Design options. The *new design* window appears:



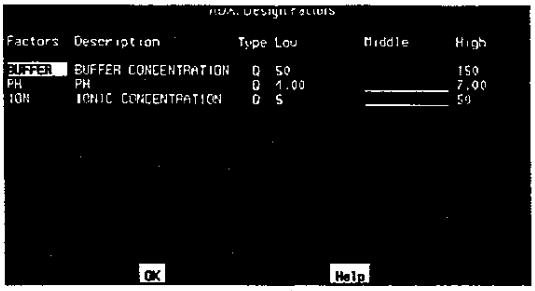
In this window each of the highlighted fields is required and must be defined. Otherwise, an experimental design cannot be generated.

First, select Type. The following window appears:



Press the Response Surface button.

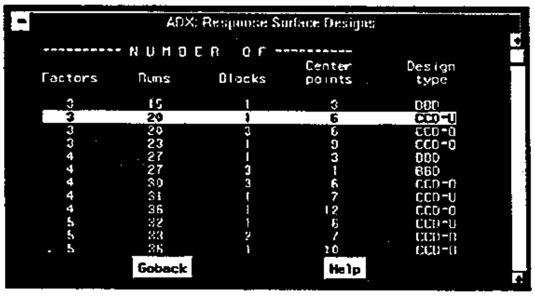
Select Factors for the next window:



The factors are defined in this window by entering the variable names and the low and high levels to consider for each variable. The middle value is automatically assigned by ADX. Press OK.

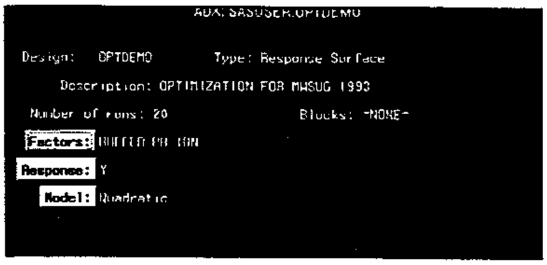
The response(s) is defined in a similar manner by selecting Responses.

Before the model is specified, the design should be selected since each design has its own corresponding mathematical models. A list of appropriate designs is generated by selecting List of available designs under the Help menu options in the *new design* window:



Choose the Central Composite design with uniform precision (CCD-U) that has twenty runs, one block, and six center points.

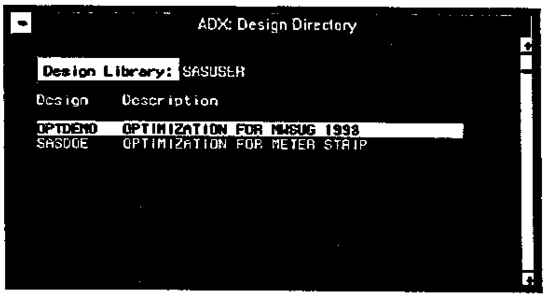
Next, since most response surfaces are quadratic, select Quadratic as the model. This will revise the *new design* window as follows:



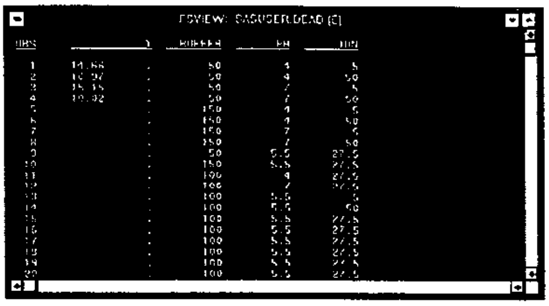
At this point, the design has been chosen but not been generated. The axial values are changed so that the suggested design space does not exceed the extreme values of the factors. This is done by selecting Parameters found in the Examine menu option and changing the Axial extreme value to 1.00. Another option, which was not implemented in this example, would have been to select Inscribe in the same window.

Finally, the design is given the name OPTDEMO.

The completed design can be generated by selecting the Save design option under the File menu. The design is then saved and displayed in the output window. To return to the *design directory* window, select End in the File menu option of the output window. OPTDEMO should appear in the window as shown below:



EDITING THE DESIGN.
By selecting OPTDEMO in the directory, the design can be edited and response values may be entered in the analysis window. Select Edit from the Modify menu option and begin to enter in the response values as follows:

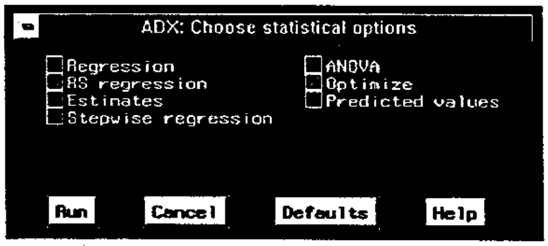


Once all the values have been entered, chose End in the File menu option to save the data.

DATA ANALYSIS.

Output.

The data can be analyzed once the data has been saved and the analysis window appears. By choosing Statistics under the Analyze menu option, the statistical options available appear in a window:



RS regression and Optimize are the default options for a response surface design. Press Run to receive the output. Because there is a great deal of SAS output, only the key sections are listed below:

****

09:06 Thursday, June 17, 1993 91

Response Surface for Variable Y: RESPONSE

Response Mean	13.162500
Root MSE	0.736138
R-Square	0.9133
Coef. of Variation	5.5927

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	3	15.131980	0.2422	9.308	0.0030
Quadratic	3	23.031052	0.3686	14.167	0.0006
Crossproduct	3	18.903750	0.3025	11.628	0.0013
Total Regress	9	57.066782	0.9133	11.701	0.0003

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Para=0	Prob > |T|
INTERCEPT	1	46.669338	5.742553	8.127	0.0000
BUFFER	1	0.010344	0.041080	0.252	0.8063
PH	1	-13.424852	2.213441	-6.065	0.0001
ION	1	0.146037	0.069043	2.115	0.0605
BUFFER*BUFFER	1	-0.000332	0.000178	-1.872	0.0907
PH*BUFFER	1	0.007367	0.003470	2.123	0.0597
PH*PH	1	1.219596	0.197292	6.182	0.0001
ION*BUFFER	1	0.001144	0.000231	4.947	0.0006
ION*PH	1	-0.018741	0.007712	-2.430	0.0354
ION*ION	1	-0.003577	0.000877	-4.079	0.0022

Parameter	Parameter Estimate From Coded Data
INTERCEPT	13.111364
BUFFER	0.793000
PH	0.318000
ION	-0.885000
BUFFER*BUFFER	-0.830909
PH*BUFFER	0.552500
PH*PH	2.744091
ION*BUFFER	1.287500
ION*PH	-0.637500
ION*ION	-1.810909

Ridge of Optimum Response

Type of ridge=MINIMUM

BUFFER	PH	ION	_PRED_	_STDERR_
100.000	5.50000	27.5000	13.1114	0.25307
96.898	5.48359	29.2473	12.9709	0.25318
94.066	5.48544	31.1155	12.7908	0.25373
91.413	5.49255	33.0333	12.5688	0.25541
88.870	5.50180	34.9776	12.3043	0.25938
86.399	5.51213	36.9377	11.9968	0.26723
83.978	5.52307	38.9081	11.6464	0.28071
81.592	5.53438	40.8856	11.2528	0.30150
79.231	5.54594	42.8682	10.8161	0.33085
76.891	5.55767	44.8544	10.3362	0.36946
74.566	5.56952	46.8435	9.8131	0.41753

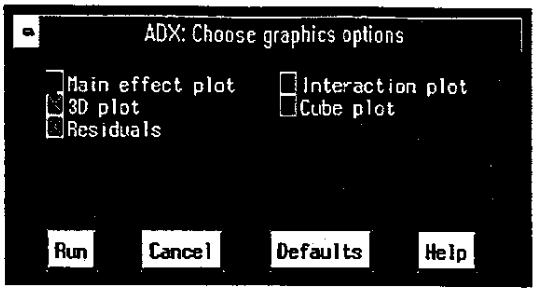
Type of ridge=MAXIMUM

BUFFER	PH	ION	_PRED_	_STDERR_
100.000	5.50000	27.5000	13.1114	0.25307
102.816	5.59199	26.2542	13.2267	0.25317
104.083	5.74877	25.7824	13.3633	0.25372
104.758	5.90612	25.5334	13.5486	0.25587
105.265	6.06067	25.3420	13.7875	0.26169
105.704	6.21356	25.1735	14.0813	0.27383
106.107	6.36546	25.0164	14.4303	0.29494
106.489	6.51674	24.8659	14.8346	0.32712
106.859	6.66760	24.7195	15.2944	0.37146
107.219	6.81818	24.5759	15.8098	0.42818
107.573	6.96854	24.4342	16.3806	0.49695

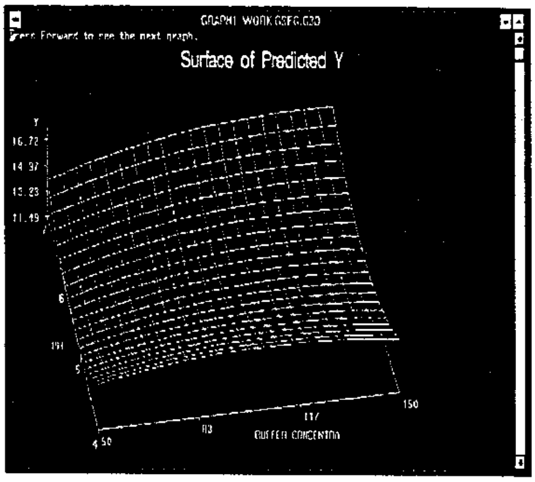
SAS 6.08

****

High resolution graphs can also be created in ADX. To graph the response surface and the residual plots, select Plots also under the Analyze menu option. The following window appears:



The 3D plot and Residuals are the default options for the response surface design. The response surface generated for this example is shown below:



Interpretation.

The sections of the SAS output (listed above) are the statistics which describe the adequacy of the model chosen, the importance of each factor in the model, as well as where the optimization of the model occurs.

The analysis produced the following predictive model for the response y:

$$
\begin{aligned}
y = {} & 46.6693 + 0.0103 * buffer - 13.4249 * pH \\
& + 0.1460 * ion + 0.0074 * buffer * pH \\
& + 0.0011 * buffer * ion - 0.0187 * pH * ion \\
& - 0.0003 * buffer^2 + 1.2200 * pH^2 \\
& - 0.0036 * ion^2
\end{aligned}
$$

It is clear that the quadratic model fits the data very well given an R^2 of 0.9133. Also, the linear, quadratic, and crossproduct terms in the model are shown to be highly significant since they all have p-values less than 0.05.

The statistically significant terms in the model are PH, PH*PH, ION*BUFFER, ION*PH, and ION*ION. The parameter estimates from the coded data show that PH*PH, ION*BUFFER, and ION*ION have the greatest effect on the mean response.

Finally, the ridge of optimum response shows at what factor levels the minimum and maximum values of the predicted response occur:

Optimum	(BUFFER,PH,ION) levels	Pred. Y
Minimum	(74.6, 5.6, 46.8)	9.8
Maximum	(107.6, 7.0, 24.4)	16.4

REFERENCES

Box G., Hunter W., and Hunter J. (1978), *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.

Brown, J., and Tobias, R. (1991), *ADX Menu System Examples*, SAS Institute Inc., Cary, NC.

Schmidt, S., and Launsby, R. (1992), *Understanding Industrial Designed Experiments* (3rd edition), Air Academy Press, Colorado Springs, CO.