

The SAS® ADX Interface – Designing the Future

Stewart L. Fossceco, Knoll Pharmaceutical Company, Whippany, NJ

Abstract

The new SAS ADX Interface provides scientists, engineers, and statisticians with the power to explore, visualize, and model a process through designed experiments. Through the visualization provided by this graphical user interface, you gain valuable insight into the behavior of a process. Scientists convey this knowledge to colleagues and management in a manner that provides the ability to make sound scientific decisions as well as efficient business decisions. This paper describes features of the ADX Interface via a pharmaceutical industry formulation case study where six factors (fluid amount, mixing time, screen size, extruder speed, disk speed, and residence time) are examined using a two-level factorial experimental design. The scientist's objectives are to identify factor levels that (1) maximize yield and (2) provide a robust process.

Key words: Experimental Design, Optimization, Statistical Graphics, Robust Process Development, Information Visualization

Introduction

Response Surface Methodology (RSM) and mixture experiments are statistical and graphical tools that help an individual to design, analyze, visualize and model a process. Typical objectives of interest are:

- Determine what factors influence a process.
- Determine how to use process factors to minimize variability in a response(s) of interest.
- Optimize the process (e.g., simultaneously maximize product yield, while minimizing product variability).

The ADX Interface for the Design and Analysis of Experiments (ADX Interface) has as its analytical foundation a series of newly developed and powerful self-contained design and analysis procedures. The ADX Interface provides an intuitive and powerful GUI to these procedures and the capabilities of the SAS System in this area. More advanced users can navigate their way within the ADX Interface to the FACTEX and OPTEX procedures of SAS/QC® to build user-defined models. By using RSM and the ADX Interface, the amount of time and the resources necessary to explore and optimize a process is reduced when compared to a “one-factor-at-a-time” experimental approach. Accordingly, the

knowledge about the process is dramatically increased compared to a one-factor-at-a-time experimental approach.

The pharmaceutical industry is charged with the task of discovering, developing, and manufacturing drug products of consistent high quality for the treatment of animal and human diseases. During the development phase of a drug, a team of individuals is devoted to the task of creating a formulation of the drug to be used in various pharmacokinetic, toxicology, and drug stability studies. These formulations can take many different forms (e.g., liquid, tablet, capsule) and are developed with specific target delivery systems in mind (e.g., oral, i.v., peritoneal, epidermal), which depend on the disease and necessary treatment regime.

Each step in this development process generates enormous amounts of data, often with confusing and sometimes conflicting results. Although any two formulation processes may differ, the one consistent theme is that individuals need the ability to make sound scientific and business decisions based on data that are properly collected, analyzed, and lucidly presented. This paper presents a case study of how the ADX Interface provides pharmaceutical scientists and statisticians the ability to design, model, and visualize a complicated process through a designed experiment, to optimize the tableted formulation, and then present these findings to colleagues and management.

The ADX Interface is available for several operating systems for Version 7 and Release 6.12 of the SAS System. The ADX Interface can be invoked in Version 7 and Release 6.12 of the SAS System by entering “xadx” at the command line. Requirements for running the ADX Interface are base SAS, SAS/QC®, SAS/GRAPH®, SAS/FSP®, and SAS/ACCESS® (necessary to import data files). For additional information about ADX Interface requirements go to <http://www.sas.com/rnd/app/qc/newadx/newadx.html>.

Tablet Formulation Background

When formulating a drug product, pharmaceutical scientists mix together several components called excipients (e.g., water, avicel, magnesium stearate, lactose, starch, etc.) with the active ingredient(s) and process them in a specific manner. The tableting process presented in this case study consists of four major stages (Figure 1), which may be part of a tablet or capsule manufacturing process:

- **Mixing** – Various excipients are mixed along with the active ingredient(s) to form a dough-like material. In this step of the process, the formulator has the option of several different machines that use different methods to mix the materials. Factors such as the duration and speed of mixing can be changed as well. Additionally, the amount of each excipient and the time it is added to the “dough” can be altered.
- **Extruding** – The dough is passed through an extruder, which makes long spaghetti-like strands. A screen with holes of a specific size determines the diameter of the extruded material. The size of the holes and the speed at which the material is pushed through the screen can be adjusted to obtain a desired type of “spaghetti” strand.
- **Spheronizing** – As the strands of “spaghetti” come out of the screen they are dropped into a bowl with a bottom like a cheese grater. This spheronizing plate can have different groove designs and can be spun at different speeds. The spheronizing plate is spun to create small pellets or beads. These pellets then go through a few additional steps such as additional drying before being compressed or coated and encapsulated.
- **Compression/Encapsulation** – The pellets are either compressed into a tablet or used to fill capsules. The compression and encapsulation processes have several steps that can be altered, but these were held constant in this case study.

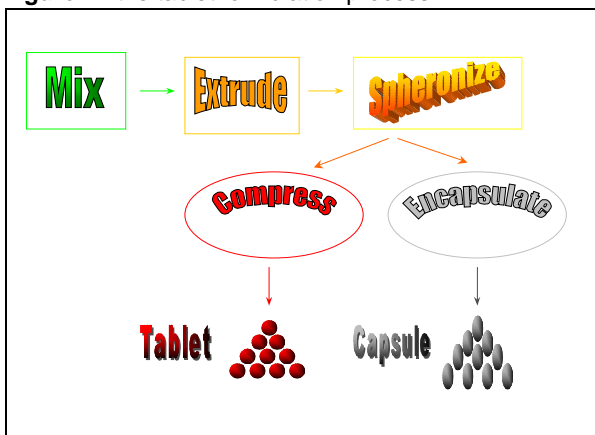
the raw levels (Myers and Montgomery, 1995); centered at 0, and scaled so that the high and low representations of the raw factor levels equal +1 and –1, respectively. This centering and scaling provides the ability to compare directly the size of effect and regression coefficient estimates. Therefore, you identify important model effects using various graphical tools that have a statistical basis for making comparisons.

Moisture content and mixing time are factors imposed during the mixing process. A specific amount of moisture in the dough relative to the current level was used (90 or 115 %), and the dough was mixed for a period of time (15 or 45 minutes). The dough was then run through an extruder at a set speed (36 or 132 rpm) using a screen with small or large holes (0.6 or 1.0 mm). Finally, the newly formed strands of material were spheronized for a specified time (2 or 5 minutes) at a set disk speed (450 or 900 rpm).

Table 1. coded and uncoded design factor levels

Factor Name	Description	-1	+1
FLUID	Moisture content, percent	90	115
MIX	Mixing time, minutes	15	45
EXTUDER	Extruder speed, rpm	36	132
SCREEN	Screen size, mm	0.6	1.0
RESID	Residence time, minutes	2	5
DISK	Disk speed, rpm	450	900

Figure 1. the tablet formulation process



The Design

The percent yield of the process that produces tablets is influenced by many factors at each stage of the process. For this case study, the scientists (Kresivic et al., 1992) chose to investigate six experimental factors, each at a high (+1) and a low (–1) level (Table 1), which resulted in a 2^6 factorial experiment (Myers and Montgomery, 1995). These ± 1 levels are the centered and scaled version of

Although there were many potentially informative responses, four were measured, and two of the four are presented here:

- Percent Yield, (YIELD)
- Mean Particle Size, (MPS).

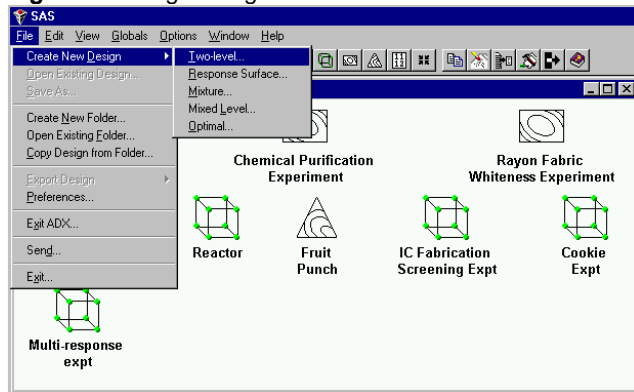
The scientists wanted to create a robust process while maximizing YIELD, and they felt that a mean particle size of roughly 1000 microns would be a characteristic of conditions that maximized yield and delivered a robust formulated product.

Using The ADX Interface

An individual intuitively navigates through the ADX Interface's desktop and screens to open, construct, or augment new folders and designs (Fulenwider, et al., 1998). Selecting “File” -> “Create New Design” from the main screen (Figure 2), the ADX Interface provides you with a simple list of design categories: “Two-level...”, “Response Surface...”, “Mixture...”, “Mixed Level...”, and “Optimal...”.

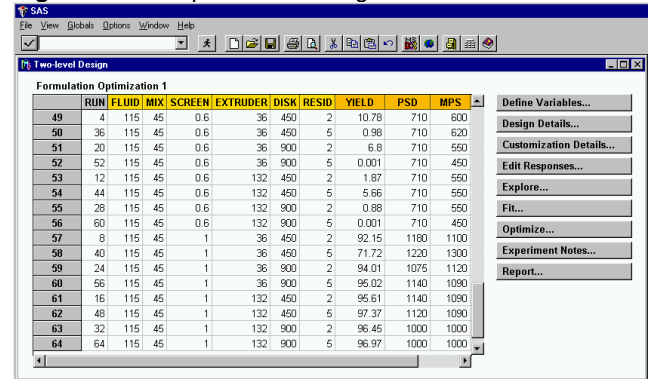
These design types can be used for specific purposes in modeling and optimizing a process. Each design type is represented graphically by a unique icon within the ADX Desktop (Figure 2: e.g., two-level by a cube, "Cookie Expt"; response surface by a contour plot in a box, "Rayon Fabric Whiteness Experiment"; mixture by a contour plot in a triangle, "Fruit Punch").

Figure 2. design categories



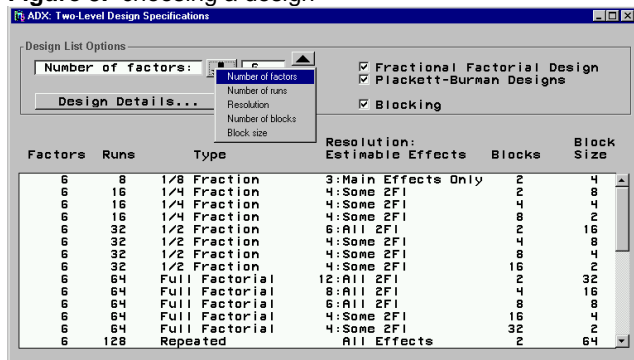
contained in "Define Variables . . .". The ADX Interface allows you to toggle easily between coded and uncoded views. Once this information and the responses are entered (via the "Edit Responses . . ." window), the data are ready to be analyzed.

Figure 4. the experimental design



Having chosen a design category, you then pick from a list of designs provided by the ADX Interface (Figure 3). You can view various design lists based on the "Number of factors", "Numbers of runs", "Resolution", "Number of blocks", or "Block size". Alternatively, you can define your own design through the "Create" -> "User Design" window. The selected design is saved to an ADX catalog, which can then be mailed to other investigators. The design can also be printed and given to the individuals performing the experiment.

Figure 3. choosing a design



In this case study, the scientists chose to run a full 2⁶ factorial, of which 16 runs are illustrated in Figure 4. The uncoded levels of design factors are entered in a table

Visualization and Analysis

It is important to understand what a model effect and interaction mean in this experimental context before beginning the analysis. In this experimental context, namely a two-level factorial experiment, an effect is very simply the difference between the average of the observed responses at the high and low levels of an experimental factor [i.e., average of $Y_{(i+1)}$ - average of $Y_{(i-1)}$]. An interaction occurs when the average response of interest at the high and low levels of an experimental factor, say A, does not behave similarly when observed across the high and low levels of another factor, say B.

Analysis involves iterating through a series of graphical views such as scatter, interaction, normal, Pareto, and factorial plots and using various model fits and checks to identify important model effects. If the important factors were included in the experiment at appropriate levels, this iterative analysis allows the scientist to arrive at a model that can be used to optimize the process. Using the ADX Interface, the analysis proceeds as follows.

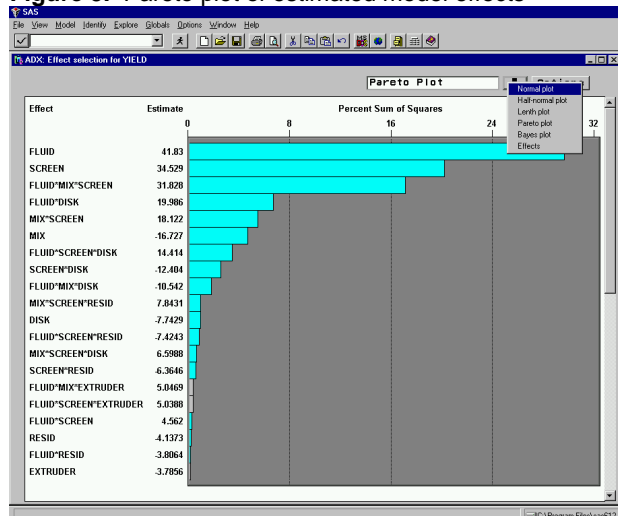
Step One: Select the "Fit..." window to fit a user-defined "Master Model" to a response with the goal of picking out the most important factors and interactions in the process for that response. In this case study, the "Master Model" included all main effects, and two and three-factor interactions. All interactions having more than three factors were insignificant and used to obtain an estimate of experimental error.

Step Two: Select the "Check Fit Assumptions" window to check assumptions required by the analysis. This window automatically appears if any model assumptions

are violated, but it can also be brought up at any time in the analysis by selecting “*Model*” -> “*Check assumptions* . . .”. When in this window, the user is guided through checks for “*Outliers*”, “*Influential Obs.*”, “*Response Transform*”, and “*Residual Plots*” by a tab format. Drill-down capabilities allow for identification of modeling and data anomalies. In this case study, several influential observations were found. These influential observations generally represented failures of the process. A graphical residual analysis did not indicate they were unusual within the context of the experiment or process and they were crucial to the analysis. A Box-Cox analysis suggested a $y^{0.4}$ data transformation, but the raw scale was determined to be sufficient and was more parsimonious. Consequently, all data were retained in the analyses and the raw observed values were used in the modeling process.

Step Three: Use the “*Effect selection*” window that comes up automatically when fitting a model and select terms to be included in the predictive model. You can define selection criterion that automatically identifies significant model terms. The ADX Interface then uses different colors, which you can define, to indicate significant and non-significant model terms. In addition to a visual representation of estimated model “*Effects*”, multiple graphical views (“*Normal*”, “*Half-normal*”, “*Length*”, “*Pareto*”, and “*Bayes*” plots) are available that make identification of significant model terms extremely intuitive (Figure 5). The Pareto principle (namely > 80 % of the work accomplished in any system is performed by 20 % of the factors influencing it) can be visualized in a plot where the absolute value of effects are ranked largest to smallest, and plotted against, say, percent sum of squares (indicator of effect importance). In this case study, there are 64 possible effects. The 20 largest (in magnitude) estimated model effects are presented in the Pareto plot (Figure 5), where the 13 largest represent the important 20 % in Pareto’s principle.

Figure 5. Pareto plot of estimated model effects

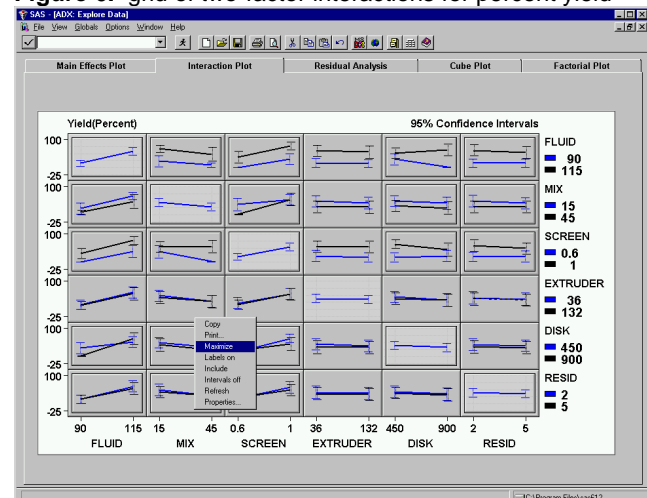


Additional views and plots indicated that these 13 model effects, together with an additional 7 model effects (main effects and interactions), were driving the YIELD and deserved further exploration.

Step Four: Select the “*Explore Data*” window to explore further the impact of main effects, and two and three-factor interactions. You can get to the “*Explore Data*” window by clicking on “*Explore*” -> “*Explore data . . .*”. Once in this window you have invaluable “*Main Effects Plot*”, “*Interaction Plot*”, “*Residual Analysis*”, “*Cube Plot*”, and “*Factorial Plot*” graphics at your fingertips.

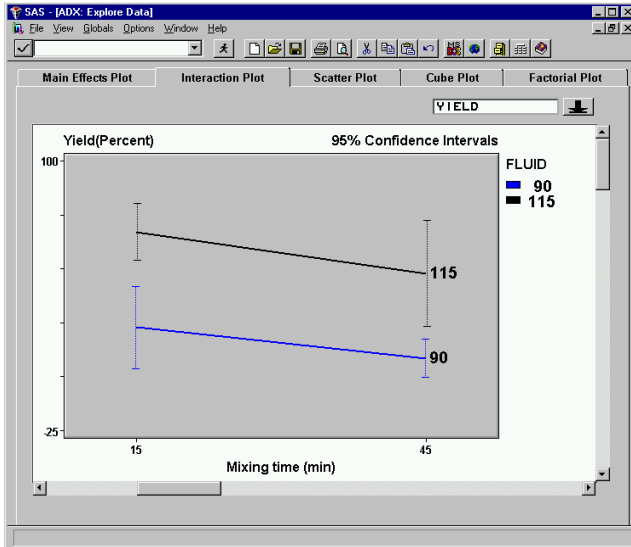
In the grid of two-factor interactions (Figure 6), you are immediately drawn to panels that are slightly raised, which indicate significant model terms. Different colored lines represent experimental high (+1) and low (-1) levels for factors on the vertical axis. The low and high factor levels for each experimental factor are listed from left to right, respectively on the horizontal axis. Vertical bars denote confidence intervals.

Figure 6. grid of two-factor interactions for percent yield



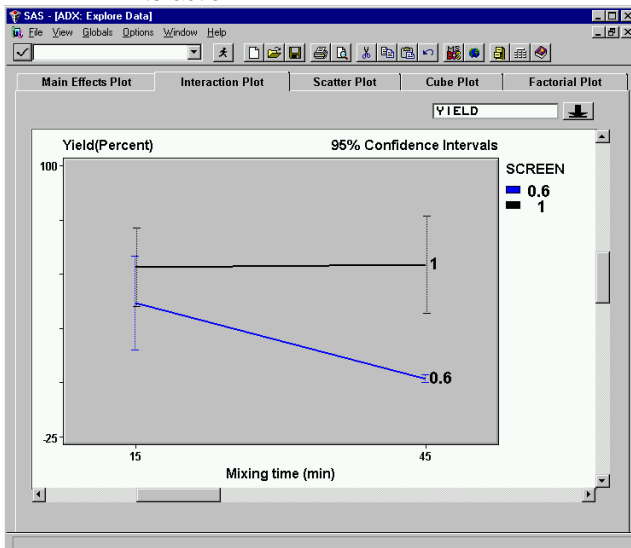
You can drill down to a single two-factor interaction of interest by using the “*Maximize*” feature with the right mouse button. In Step Three the two-factor interaction of FLUID*MIX was not found to be significant, a fact more saliently revealed by parallel lines in the maximized FLUID*MIX two-factor interaction frame in Figure 7. Thus, the effect the level of MIX has on YIELD appears not to be influenced by the level of FLUID.

Figure 7. maximized FLUID*MIX two-factor interaction



In contrast, the SCREEN*MIX two-factor interaction was found to be significant in Step Three, which is apparent by the non-parallel lines in Figure 8. Notice that the YIELD stays constant for any mixing time (level of MIX) as long the material is extruded through a large screen (1.0 mm); however, YIELD drops precipitously (near 0 %) if the material is mixed for 45 minutes and then extruded through a small screen (0.6 mm).

Figure 8. maximized SCREEN*MIX two-factor interaction

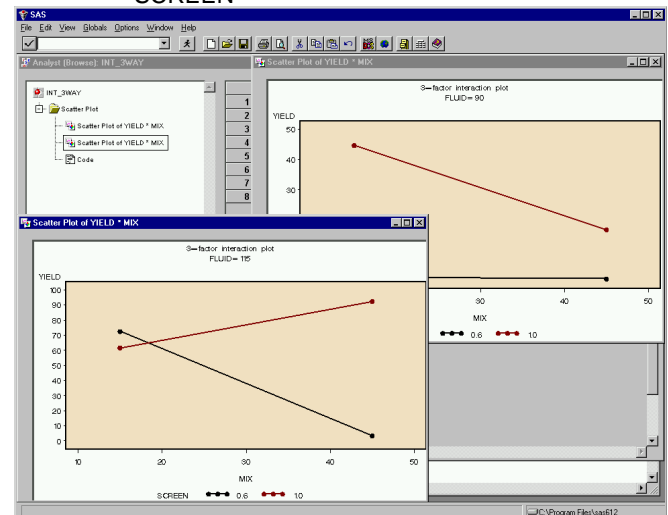


The investigation does not end here. Again, recall that in Step Three the three-factor interaction FLUID*MIX*SCREEN was found to be significant (the

third largest effect, in magnitude, Figure 5). Figure 9 is a graphical representation of this three-factor interaction. Note that the average YIELD is almost 100% for FLUID = 115 %, MIX = 45 minutes, and SCREEN = 1.0 mm. Conversely, the average YIELD is approximately 20 % for FLUID = 90 %, MIX = 45 minutes, and SCREEN = 1.0 mm, and there was almost no YIELD at all when FLUID = 90 % and SCREEN = 0.6 mm, regardless of the mixing time. Hence, the graphical representation of the FLUID*MIX*SCREEN interaction (obtained using the Analyst Application) in Figure 9 illustrates several important points to the scientist:

- The SCREEN*MIX two-factor interaction plot at the high level of FLUID is not the same as the two-factor interaction plot at the low level of FLUID. This by definition is what a three-factor interaction looks like. This implies that FLUID and MIX do interact, which is contrary to what was concluded in Step Three. The expression of the FLUID*MIX two-factor interaction, however, depends on the level of SCREEN. As shown here, two-factor interactions can be very deceiving in the presence of a significant three-factor interaction. Not identifying and including a significant three-factor interaction could result in wrong conclusions.
- A huge difference in scale exists between the FLUID=115 and 90 plots. FLUID is very important.
- Having low levels of both SCREEN and FLUID is a very bad choice of operating conditions.
- If a higher level of FLUID is chosen, there is more flexibility in the choice of where to operate for both MIX and SCREEN. That is, choosing the high level of FLUID results in a product that is robust (YIELD remains high) to the level of MIX and SCREEN within the explored design space.

Figure 9. three-factor interaction plot for FLUID, MIX and SCREEN



Additional exploration further revealed the extremely important relationships (interactions) among FLUID, MIX, SCREEN, DISK, and RESID and their impact on YIELD.

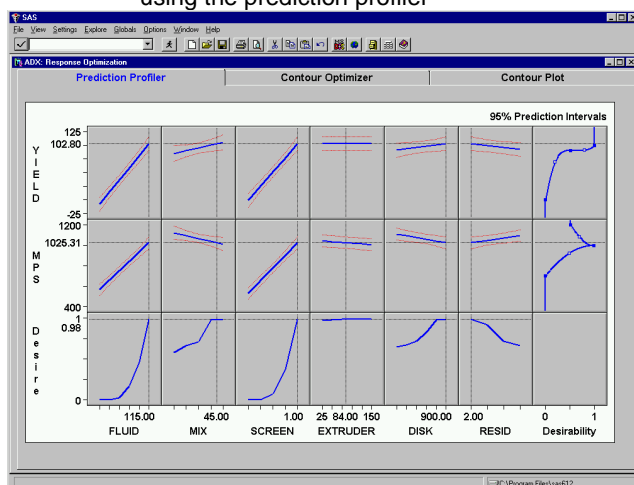
After iterating through this process of modeling, checking, and graphical analysis, models were chosen for YIELD and MPS. In the ADX Interface, these models can be, but are not necessarily, identical. In this case study, the final predictive models chosen for YIELD and MPS were in fact not the same. In the process of choosing a final predictive model for each response, a great deal of knowledge was gained about how the product was behaving in the process, and a story was beginning to come together about the product's behavior.

- FLUID, MIX and SCREEN are very influential in the process.
- There is a very complex relationship among the experimental factors in this process, as is evident through the number of important two and three-factor interactions, for example, FLUID*MIX*SCREEN, FLUID*SCREEN*DISK, MIX*SCREEN*RESID, . . .
- DISK and RESID impact the process mainly through two and three-factor interactions with FLUID, MIX, and SCREEN.
- Within the design space, EXTRUDER did not significantly affect the system.

Because of this knowledge, models were chosen. The next step was process optimization.

Step Five: Select the "Optimize . . ." function to choose the responses to be optimized where the predictive model has already been elucidated through the "Fit . . ." and "Explore . . ." process. Once in the "Response Optimization" window (Figure 10), you have the ability to explore the effect of factor settings on the predicted responses through the "Prediction Profiler", "Contour Optimizer", and "Contour Plot" windows.

Figure 10. simultaneous optimization of YIELD and MPS using the prediction profiler



The "View -> Show Desirability" option (which when turned on provides the right hand column in Figure 10) of the "Prediction Profiler" window provides an intuitive method of simultaneously examining the effect that moving from the low to the high level of an experimental factor has on multiple responses, namely YIELD and MPS. If it is more desirable to have a response close to say a maximum, its desirability is then 1 at its maximum value, and goes to 0 at its minimum (e.g., YIELD). Similarly, if there is a targeted value, then the desirability is 1 at the target, and goes to 0 (not necessarily in a linear fashion) as the response moves away from the targeted value (e.g., MPS).

By observation (Figure 10), the optimal settings for FLUID, MIX, SCREEN, RESID, and DISK were predicted to be at 115 %, 45 min, 1.0 mm, 2.0 min, and 900 rpm, respectively with an observed mean YIELD of 96.5 % (Table 2). This is in stark contrast to what was found when interpreting previous one-factor-at-a-time experiments where the optimal settings were found to be 115 %, 15 min, 1.0 mm, 2.0 min, and 450 rpm, respectively with an observed mean YIELD of 64.3 % (Table 2). (Because EXTRUDER was found to be unimportant in the process, it can be held constant anywhere within the design space and it will not significantly impact YIELD.) These differences in the predicted optimum settings were due to the overwhelming influence of the two and three-factor interactions. You can obtain predicted responses by using the "Response Calculator" window ("Explore" -> "Response calculator...") to calculate the predicted response for desired factor level settings.

Table 2. observed and predicted percent yield with design factors at designated levels.

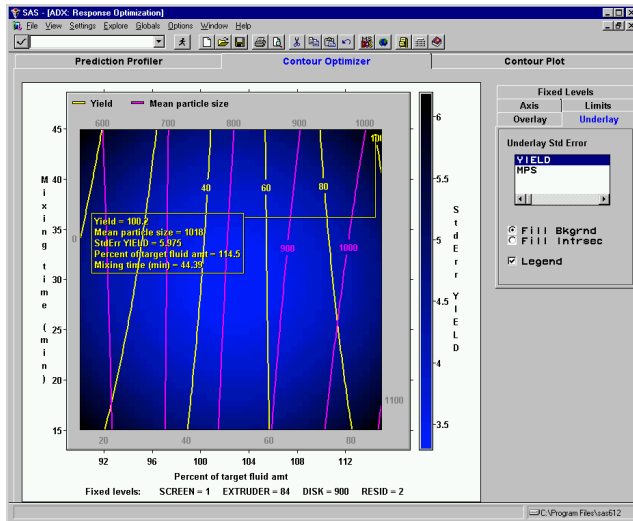
Factor	Settings (one-factor-at-a-time)	Settings (RSM)	Confirmatory Batch
Fluid amount	HIGH, 115	HIGH, 115	112
Mixing time	LOW, 15	HIGH, 45	40
Extruder speed	LOW, 36	HIGH, 132	82
Screen size	HIGH, 1.0	HIGH, 1.0	0.8
Residence time	LOW, 2	LOW, 2	3
Disk speed	LOW, 450	HIGH, 900	550
Observed Yield	64.3 %	96.5 %	99.9 %

The "Contour Optimizer" window can be used to explore further the behavior of the responses with respect to the experimental factors by overlaying contours of constant predicted mean YIELD and MPS, and underlying the standard error of YIELD (Figure 11). The optimal factor settings found in the "Prediction Profiler" were applied directly to the contour plot shown in Figure 11 by selecting "Settings" -> "Apply settings from Profiler".

By using the right mouse button, additional overlay levels can be added to the graph to highlight interesting features

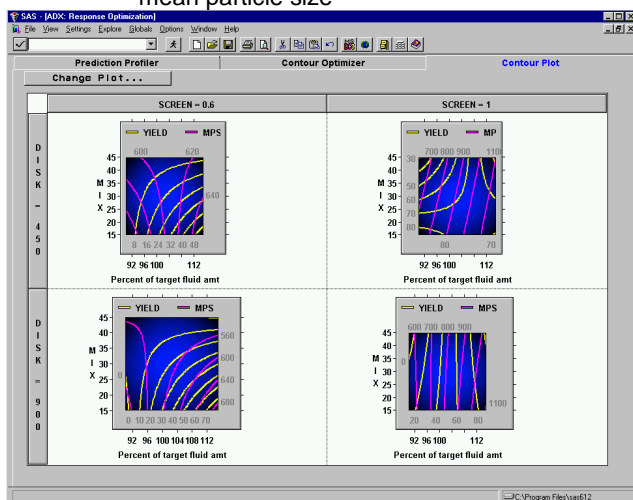
of the surface. Additionally, you can use the right mouse button to annotate the surface interactively, focusing attention to an area of special interest within the design region.

Figure 11. contour plot annotated to highlight a particular area of interest



The “Change Plot ...” functionality of the “Contour Plot” window can also be used to produce slices of contour surfaces in up to four dimensions, while holding all others constant at user-defined levels (Figure 12).

Figure 12. multiple contour slices for percent yield and mean particle size



You can see how the shape of the contours change for YIELD and MPS over different levels of FLUID and MIX as you move from a SCREEN of 0.6 to 1.0 mm and a

DISK of 450 to 900 rpm. This set of contour plots clearly illustrates how interactions among the experimental factors dramatically change the shape of the response surfaces for YIELD and MPS as the factor levels change from their high to low levels. Figure 11 (a region of particular interest) is, in fact, the lower right hand pane of Figure 12.

Discussion

The optimal factor settings and observed yields found using RMS methods with the ADX Interface differ substantially from the original recommended settings that were based on a one-factor-at-a-time experimental approach (Table 2).

The reason a one-factor-at-a-time approach failed to identify properly an optimal region in which to operate was the existence of overwhelming significant two and three-factor interactions. The FLUID, SCREEN, MIX, RESID, and DISK were all involved in very significant two and three-factor interactions. Interaction effects are generally difficult to identify, and they are even harder to identify when factors involved in the two and three-factor interactions are from different steps in the process. FLUID and MIX are factors from the very first step in the process, whereas SCREEN is a factor from the second step, and RESID and DISK are factors from the third step.

Some scientists initially believed that the factors studied in this experiment only impacted YIELD at the time the factors entered the process; for example, FLUID only influenced the process during the mixing of the excipients, and had no influence at the extrusion phase. This case study clearly demonstrates that factors located anywhere in a process can have an impact at many different places through interactions with other factor(s) from seemingly unrelated locations in that process.

All batches for this experiment were prepared in a laboratory on small-scale equipment to conserve drug and save time. The confirmatory batch (Table 2) was the only full-scale batch made, and was used to confirm that the settings defined based on the analysis of these experimental data maximized yield.

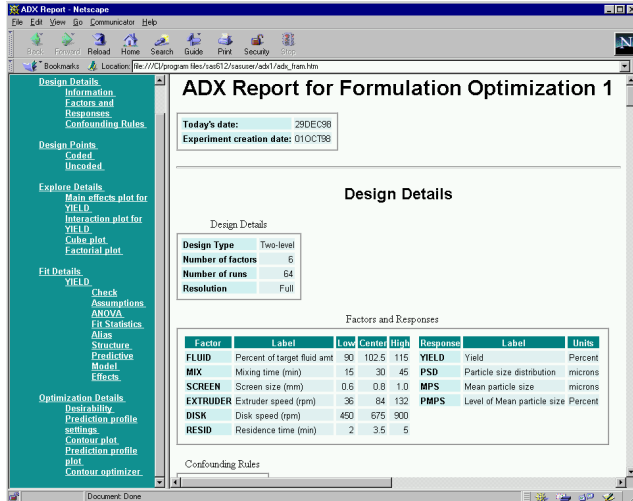
Communicating Findings

Producing documentation of analytical results can be extremely time-consuming. The ADX Interface provides a reporting tool that generates reports in two forms. These two forms expedite the documentation process and enhance written communication.

The first reporting form is a self-generating html formatted report that includes user-chosen graphics, equations, and details about the design and analyses (Figure 13). This

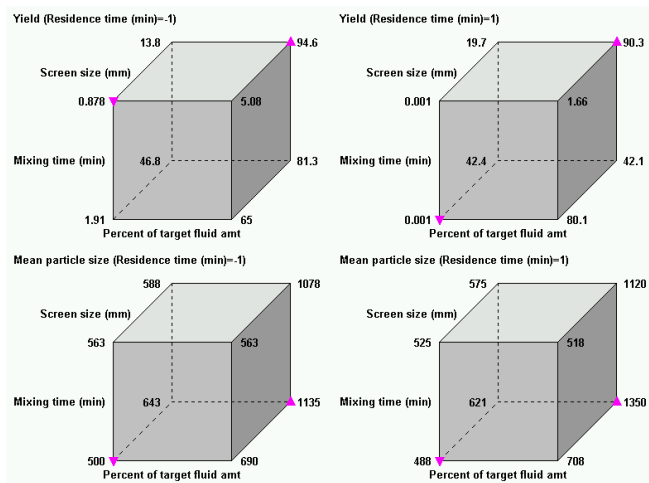
report is then viewed using an internet browser, and can be published on a company intranet.

Figure 13. sample html output from an ADX Interface generated html report using the Netscape browser as the viewer



The second reporting form generates analytical results (e.g., regression equations, optimal settings, etc.) in a text-formatted file. No graphs are generated in this text-formatted file. However, graphics are exported easily using one of several graphics formats (e.g., BMP, GIF, TIF, WMF, etc.), or copied and pasted into a document using the clipboard (Figure 14). You then combine the text and graphics using a word processor into a single report.

Figure 14. sample graphics output copied to and pasted from the clipboard of the ADX Interface



Summary

Designed experiments are valuable tools for research and development, and quality improvement initiatives in many industries. These designs are often complex to construct, and historically a statistician has performed a substantial amount of programming to design the experiments and analyze the data. The amount of time necessary to design and complete a thorough analysis of all designed experiments by a statistician is not always possible in today's business environment. The ADX Interface provides statisticians and scientists the ability to expedite the design and analysis of many experiments all within a GUI. The thoroughness of the analysis does not suffer from the acceleration of these efforts, but alternatively it is enhanced due to the ADX Interface. In fact, more time is available for pre-design discussion and investigation. Additional time is also made available to construct designs and perform analyses that are more intricate because of the time saved by using the ADX Interface for the design, visualization, and analysis of more commonly used experimental designs. These additional efforts enhance the possibility of a successful outcome.

The outcome of analyzing the data collected for this designed experiment was a business decision. After analyzing the data collected in this case study, the formulators presented their findings to the project team. The team manufactured a full-scale production batch to confirm the experimental findings. This confirmatory batch had an observed YIELD of 99.9 %. The team decided to shift production to the "optimal" region, centered at the factor settings where the confirmatory batch was taken. The improved production process, which included redefined factor limits, was validated by using the small-scale work presented in this case study. Percent yield has since been over 99 %, and to the best knowledge of the author, a batch of material has never failed.

Acknowledgements

I would like to thank the following individuals for their insightful contributions:

- Donna Fulenwider, SAS Institute Inc.
- Fred Immermann, Wyeth-Ayerst Research
- John Kresivic, Wyeth-Ayerst Research
- Maura Stokes, SAS Institute Inc.

References

Fulenwider, D. O., R. D. Tobias, R. N. Rodriguez. (1998), "SAS® System Tools for Design and Analysis of Experiments," *Proceedings of the Twenty-Third Annual SAS Users Group International Conference*, 23, 1172-1182.

Krešević, J., N. Sheth, W. Cloud and M. Ganesan. (1992), "The Effect of Extrusion/Spheronization Manufacturing Variables on Acetazolamide Pellets," *Pharmaceutical Research October Supplement*, 9:10, p S-117, ISSN 0724-8741, Plenum Press: New York and London.

Myers, R. H. and D. Montgomery. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: John Wiley & Sons.

SAS, SAS/ACCESS, SAS/FSP, SAS/GRAPH, and SAS/QC software are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Author

Stewart L. Fossceco, Knoll Pharmaceutical Company, North Jefferson Road, Whippany, NJ 07981, Phone: 973-428-4140, Email: fossces@basf.com; 46 Center Grove Road, Building U #502, Randolph, NJ 07869, Phone: 973-328-3760, Email: fossceco@mindspring.com.