CRITICAL PROCESS PARAMETER DETERMINATION DURING PRODUCTION START-UP

by

Christine M. Lindsey

B.S. Chemical Engineering, Rice University, 1999

Submitted to the Sloan School of Management and the Department of Chemical Engineering in partial fulfillment of the requirements for the degrees of

Master of Business Administration
AND
Master of Science in Chemical Engineering

In conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology
June 2004

Signature of Author ____
May 6, 2004
MIT Sloan School of Management
Department of Chemical Engineering

Certified by ____
Roy E. Welsch, Thesis Advisor
Professor of Statistics and Management Science
Center for Computational Research in Economics and Management Scince

Certified by ____
Gregory J. McRae, Thesis Advisor
Bayer Professor of Chemical Engineering

Accepted by ____
Margaret Andrews
Director of Masters Program
Sloan School of Management

Accepted by ____
~~Daniel Blankschtein~~
Chairman, Committee for Graduate Studies
Department of Chemical Engineering

CRITICAL PROCESS PARAMETER DETERMINATION DURING PRODUCTION START-UP

by

Christine M. Lindsey

Submitted to the Sloan School of Management
and the Department of Chemical Engineering
on May 7, 2004
in Partial Fulfillment of the Requirements for the Degrees of
Master of Business Administration
and
Master of Science in Chemical Engineering

Abstract

Production start-up data is consistently utilized in a reactive manner during the initial stages of a product's lifecycle. However, if proactive information systems are created before full scale production starts, ramp-up cycles can be shortened considerably. This project attempts to develop a framework for analyzing process data quickly and efficiently during a new product start-up in order to provide information for the short term goals relating to attaining stable processes as well as provide guidance on long term handles for process improvement.

First, a summary of previous literature regarding start-up process data as well as typical stable process data usage will be presented. This will provide adequate background for evaluating typical gaps present during production ramp-up. Then, solutions to these gaps will be discussed in order to develop tools for better data analysis in shorter periods of time. These methods will then be applied to a case study involving the new production of Kodak's DCS Pro 14N digital camera.

The Kodak Professional DCS Pro 14N was an amazing leap in technology: a camera with double the resolution for roughly half the price of any product available. Unfortunately, it soon became apparent that the original demand estimates were grossly underestimated, straining original resource allocations. Manufacturing struggled to start and was already a year behind in backorders. With over 1500 process attributes collected on each camera, the key drivers of quality had yet to be determined. The surrounding circumstances made the quick analysis of start-up data vital to effective resource management and yield improvement of the camera.

After using the new process modeling framework and modified control techniques on the example Kodak case, two additional topics will be discussed. First, the many classifications of return on investment in proactive start-up data analysis will be presented. Ranging from waste minimization to higher customer satisfaction, these incentives justify early preparation for start-up data analysis. Finally, future areas of study will be recommended to augment the findings within the thesis.

This page intentionally left blank.

Acknowledgements

Without the support of many special individuals, my time at MIT and Kodak would not have been as enjoyable.

First, I would like to thank the many individuals at the Eastman Kodak Company who made my project a success during a very intense product introduction. Without the vision and support of my project supervisor, Vincent Andrews, the project scope would not have been half as ambitious as it was. The daily interaction and guidance from Gerry Edd, Alan Butenhof, Rae Mawn, Christopher Powell, Eric Dilella, Tim Keefe, and especially Eshetu Setgen gave me the strength to finish a project that I initially considered impossible. Finally, I could not have finished without the significant help of Jim Merriman, Richard Kurchyne, Marty Maurinus, Kevin DeBaise and Patricia Reibstein.

My thesis supervisors were very generous with their time and knowledge. Prof. Gregory McRae provided a wonderfully integrated approach to a challenging problem. The valuable guidance given by Professor Roy Welsch supplied practical solutions which ultimately allowed full implementation of the project's broad scope.

The two year experience in the Leaders for Manufacturing Program has been among the best in my life. I would like to thank all of the sponsor companies and the LFM staff for the exposure to so many different learning opportunities. Without Don Rosenfield, Bill Hanson, and Nancy Young the experience would not have been the same.

Beyond the two years of late nights and long papers, LFM brought together a great group of people. Throughout this project James Katzen, Elana Cohen, Kevin Schwain, Christopher Pandolfo, and Yuliya Frenkel were indispensable sounding boards. The LFM program would not be complete without every last one of the fabulous fifty seven individuals within the Class of 2004.

Finally, I would like to thank my family and my husband, Michael. Without their support and love my life would not be the same.

This page intentionally left blank.

Table of Contents - Critical Process Parameter Determination during Production Start-up

1.0 Introduction 11

1.0.1 Situation

1.0.2 Problem Statement

Figure 1: Cycle of Process Data Usages 13

2.0 Theoretical Background 15

2.1 Design Structure Matrix 16

2.1.1 Mechanics of DSM

Figure 2: Example Design Structure Matrix 17

Figure 3: Guidelines for mapping interactions 17

Figure 4: DSM Rearrangement displays clusters 18

2.1.2 Use Parameter-Based Data Type for Process Modeling

2.2 Approaches to Statistical Process Control 20

2.2.1 Quantifying Process Variability

2.2.2 Displaying Process Variability, Single Variables

2.2.3 Multivariate Process Monitoring

Figure 5: Basic Illustration of Bivariate Relationship 25

2.2.4 Traditional Short-Run SPC Solutions

2.2.5 Summary

2.3 Theoretical Background Summary 27

3.0 Data Collection and Analysis Strategies 29

3.1 Data Collection Strategies 29

3.1.1 Team Structure

Table 1: Resource allocation table enabling project assignments 30

3.1.2 Creating a Data Structure

3.1.2.1 "Firefighting" Data Structure

3.1.2.2 Proactive analysis structure

3.1.3 Summary

Figure 6: Self-reinforcing decision making cycle 34

3.2 Overall Process Data Mapping Approaches 35

3.2.1 Background

Figure 7: Diagram showing the relationship of process conditions and product attributes 36

3.2.2 Single Stage Analysis

Figure 8: Eliminating redundant information 39

3.3.3 Sequential Stage Analysis
Figure 9: A stage diagram tracks the attributes that are incoming and outgoing from each stage 42
3.3.4 Overall Process Descriptions
3.3.5 Combination of DSM and PLS methodologies
Figure 10: Example matrix representation of model and observed process variables 44
Figure 11: Results of one iteration of process variable reduction 46
Figure 12: Results of two iterations of process variable reduction 46
3.3.6 Residual Pattern Analysis
Figure 13: Original data, with 2 places behind the decimal points and data rounded to the nearest whole number 48
Figure 14: Original Data's residuals 48
Figure 15: Random error around Equation 1, Least squares fit of data 49
Figure 16: Slopes in the residuals indicate an error in the model 50
Figure 17: Random error around Equation 1, Least squares fit of data 50
Figure 18: Slopes in the residuals indicate an error in the model 51
Figure 19: Slope of the model equals the underlying model 52
3.5 Control Charting for Production Start-up 53
3.5.1 Introduction
3.5.2 Desired Attributes
Figure 20: Start-up manufacturing phases 54
3.5.3 An Interim Solution
3.5.4 Examples
Figure 21: Variable Control Chart, XBar using subsets of five 58
Figure 22: Control Chart, Individual Measurement of Variable 1 58
Figure 23: Interim Solution Control Chart 59
Figure 24: Interim control chart with dynamically calculated control limit on the first 80 data points 60
Figure 25: Interim Control Chart with Specifications, all 100 data points 61
Figure 26: Interim Control Chart with Specifications, first 80 data points 61
3.5.5 Including Robust Techniques within Interim Solution
3.5.6 Summary
4.0 Case Example: Kodak Digital Camera 65
4.1 Background 65
Figure 27: Process Outline 67
4.2 Data Generated 67
Figure 28: Diagram of data types and their related process stages 68

4.3 Data Analysis 69
Figure 29: Sub-processes within the Calibration process step 70
4.4 Determination of Critical Process Parameters Example 70
Figure 30: Scatterplot matrix (bivariate relationships) of the four final product attributes and seven in-process variables 71
Figure 31: Fit of Final Attribute 1 with Final Attribute 4 72
Figure 32: Summary Process Correlation Matrices 73
Figure 33: Reduced Process Correlation Matrices 73
Figure 34: Scatterplot of bivariate relationships present in the sequential stage analysis 74
Figure 35: Final attribute 2 and a linear combination of final attribute 1 and in-process 2 75
Figure 36: Last 100 Final Attribute Data Points using Traditional Control Limits 76
Figure 37: Last 100 Final Attribute Data Points using Capability Limits 76
Figure 38: Last 100 Final Attribute Data Points using Robust Capability Limits 77
4.5 Summary 78
5.0 Conclusions and Further Areas of Study 79

Bibliography 83
Appendix I: Detailed Model Output for Kodak Case Example 85

This page intentionally left blank.

1.0 Introduction

Today's start-up manufacturing environment is increasingly complex. As production fixtures contain more computer controlled parts, more process data is available for analysis. At the same time, the design cycles for each product are shortening to the point where complete process models are rarely completed before full-scale manufacturing begins. External factors compound these trends as the global business environment increases the number and quality of competitors. Each of these issues is further exacerbated if the product itself is complex, as measured parameters increased, development cycles are more involved, and competitors develop unique resources. These recent developments are additive to the already challenging atmosphere of a production start-up. In most cases, there are daily challenges that consume the attention of all available resources. This lowers the priority of mapping and modeling the process – both necessary precursors for establishing meaningful Statistical Process Control (SPC). Instead, development data or data from the preceding product cycle is substituted leading to many false assumptions.

This thesis attempts to outline a method for efficiently creating adequate process models and SPC tools using traditional SPC techniques, the Projection to Latent Structure method (PLS), and the Design Structure Matrix (DSM) methodology together. This combination of parameter determining tools will be exemplified using a case from the Eastman Kodak Company's recent start-up of a new professional digital camera, the DCS Pro 14N.

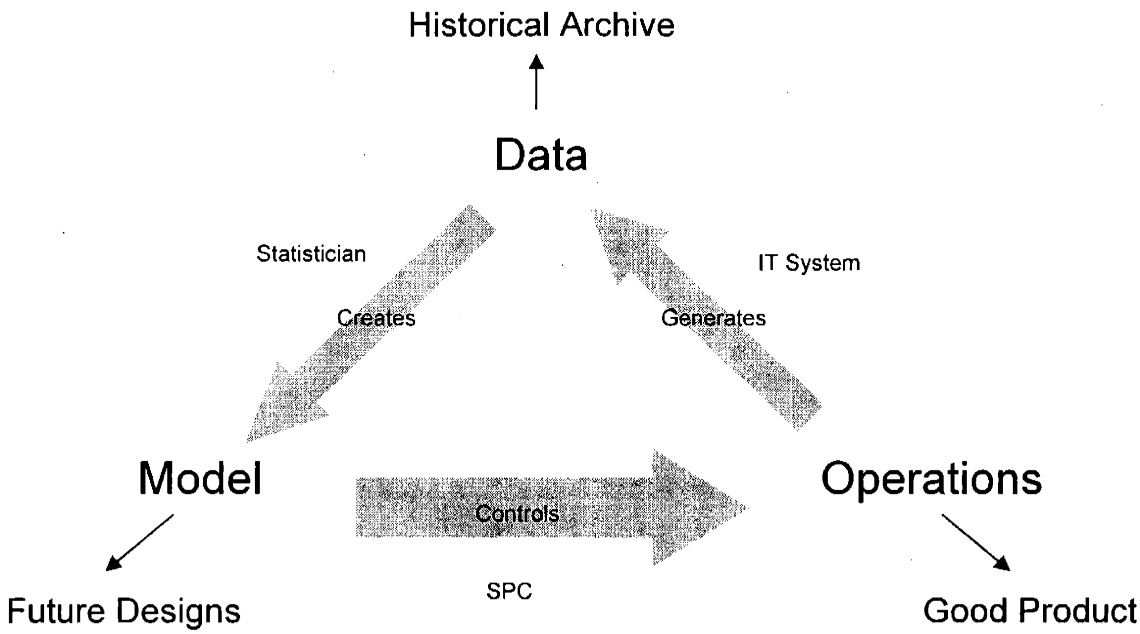
With the recent corporate attention on six sigma quality and ISO certification, production quality systems are more visible than ever before. What many individuals do not understand is that a final quality inspection is only the first step to creating consistent product. Because manufacturing processes control the outcome of product attributes, it is imperative to find the causes of final product variability as far back in the process flow as possible. Therefore, while traditional SPC can be used for the final product attributes, the full power of SPC will only be realized when applied to the process conditions which control the outcome of the final product attributes.

The "Control" part of SPC is an interesting term since SPC is mainly a passive tool for the display of data. Without fully understanding the interactions within the process and how they affect the final product attributes, an "in-control" process in SPC terms only means that the process is consistent[1]. Only when used with process modeling knowledge is the control aspect of SPC valid. Because of this dependent relationship, SPC and thorough process analysis should be coupled efforts.

In an ideal situation, the all process data would be used in a loop designed for continuous improvement of the current product itself as well as future product generations. As product is manufactured, certain data is produced. This data is captured via an Information Technology (IT) system which serves many functions from archiving the data for later analysis to display of critical information to operators within the process itself. This data can then be used by a statistician or engineer to create a quantitative process model of the overall manufacturing system. During the course of model creation, certain critical process variables will become apparent. These critical variables can be used as production targets and to ultimately control the process through SPC. And, as process variables that appear to drive certain final product attributes are changed other process relationship might surface. This begins another cycle of the data analysis which will eventually narrow to a production set of manufacturing process conditions. The following figure illustrates this regenerative cycle for process data.

[1] Memory Jogger.

Figure 1: Cycle of Process Data Usages[2]



Challenges arise when this cycle begins, as all the steps are dependent on each other. One obvious difficulty occurs when there is no previous data for the statistician to create a mathematical model of the system or when the data is found to be invalid. Additionally, some data is not of equivalent quality and relationships between variables can be perceived or more qualitative in nature. This is often the case during a production start-up and forces both the process modeling and SPC efforts to be delayed while data is collected. The length of this delay can vary based on the number of units produced, the prior knowledge of the process as determined by the development cycle, and the amount of resources accessible. In the interim, there are some options available to approximate process control further upstream. With the right type of data collection, establishing process relationships around the most critical product attributes and then applying limited control charts can be effective in the short term. As more data is collected, the cycle can continue toward the goals of lowering cost and increasing product consistency.

[2] Modified from Perry's Chemical Engineering Handbook.

This page intentionally left blank.

2.0 Theoretical Background

When analyzing data from a production start-up, there are two main challenges to overcome. First, quantifying the various process variable interactions provides a cumbersome task during the typical daily struggles. Variable interactions can range from simple, linear relationships in subsequent process steps to non-intuitive multi-step interactions. If the proper amount of attention is given to finding a descriptive process model, this will lead manufacturing to more proactive process improvements over reactive process maintenance. This section will provide an introduction to the Design Structure Matrix methodology for tracking the strength of process step interactions.

A second major challenge is providing adequate information about the process behavior to the individuals that can use it most effectively. In typical production environments, a mean, upper control limit and lower control limit are calculated from historical data for each critical production variable. These are then used as guides with the appropriate run tests to alert production staff if the process is drifting in a bad direction or has more variation than expected. Some typical Statistical Process Control (SPC) techniques will be described in the following sections in order to provide an introduction to the strengths and weaknesses of these methods for the start-up production environment.

2.1 Design Structure Matrix

Originally outlined by Donald Steward in the early 1980's to better track the design process of integrated circuits and lower the feedback iterations[3], the Design Structure Matrix (DSM) methodology provides a valuable tool to map interactions clearly. DSM (also known as the Dependency Structure Matrix and the Design Precedence Matrix) was further popularized by Steven Eppinger and Robert Smith to better describe interactions within corporate organizations, especially within the context of product design phases. While there are many new innovations to this technique, the work presented in this thesis relies on the very basics of the procedure.

Four main data types of DSM have been identified: component-based, team-based, activity-based, and parameter-based[4]. The component-based DSM tracks the interaction between various elements in a more complex system. An example of this type of data type would be a matrix showing the interaction of various subroutines within a larger computer program. The next two data-types, team-based and activity-based are focused on the interactions within organizations. While team-based analyzes the interfacing of individuals, the activity-based analysis tracks the activities of those individuals in order to better plan projects. Finally, the parameter-based data types capture the interactions between process parameters. This final category will find extensive use in the following sections.

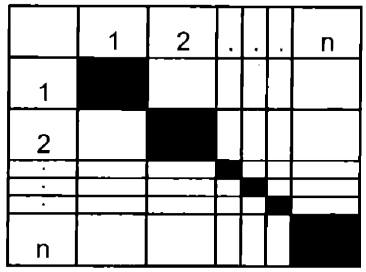
2.1.1 Mechanics of DSM

DSM, as its name suggests, is centered on the use of a matrix. This matrix is formed by taking all of the possible *n* system inputs, and creating a row and column for each respective input. In the case of product design, there might be a list of n activities that need to occur in order for the design phase to be completed. Ordered the same way in the row and column dimension, this forms and *n* by *n* matrix, exemplified below:

[3] Welch.

[4] http://www.dsmweb.org/Tutorial/DSM_types.htm

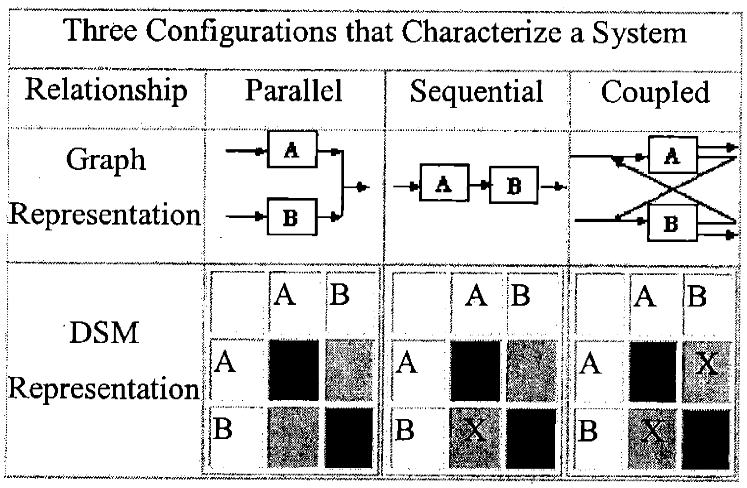
Figure 2: Example Design Structure Matrix



The cells along the diagonal are blacked out to denote the assumption that interactions within a given design step are not applicable.

The next phase in DSM methodology is to map the interactions between the steps ranging from 1 to n. To maintain consistency, guidelines for marking the various types of interactions were developed. For instance, if a process B initiation relies on the process A being finished first, an "X" is entered below the diagonal in the box shared by the row B and the column A. Similarly, if processes A and B depend on information from each other "X" entries are made in both cells common to A and B rows and columns.

Figure 3: Guidelines for mapping interactions. Source: http://www.dsmweb.org/Tutorial/tutorial_intro.htm[5]



From the above table, it is observed that the entries within the table are binary. There have been recent advances in standardizing new ways to capture additional information ranging from numerical cell entries to color coding schemes to better capture interaction strengths. While these

[5] The Design Structure Matrix Homepage hosted by MIT. www.dsmweb.org

modifications capture more detail about the system, their purpose is to act as an extension to the goal of mapping the interactions of a system of variables.

Using whatever preferred method for individual entries, the next phase in the DSM methodology revolves around rearranging the rows and columns in a way to group the interactions as closely as possible about the diagonal. As the following example illustrates using a team-based data type, this activity should result in obvious clusters.

Figure 4: Rearranging the DSM allows for obvious clusters of interactions to be highlighted. The matrix on the left denotes an initial DSM attempt while the right matrix shows the result of rearranging the rows and columns to best cluster the interactions. Source: http://www.dsmweb.org/Tutorial/clustering.htm

	1	2	3	4	5	6	7
1		X			X	X	
2				X			X
3		X		X			X
4		X	X		X		X
5				X		X	
6	X				X		
7		X	X	X			

	1	6	5	4	2	3	7
1		X	X		X		
6	X		X				
5		X		X			
4			X		X	X	X
2				X			X
3				X	X		X
7				X	X	X	

In this case, the clusters suggest teams of individuals [1, 6]; [4, 5] and [2, 3, 4, 7]. Optimizing the rearrangement of rows and columns can be done manually if there are complicating factors which warrant the additional attention. Otherwise computer algorithms exist and are available to determine the most efficient formation of teams.

2.1.3 Use Parameter-Based Data Type for Process Modeling

DSM has strength in its simplicity. This makes it an obvious candidate for tracking numerous process parameters and variables present in many of today's manufacturing environments. By simplifying the process modeling process using DSM, faster analysis and therefore improvements can be made. This can be especially beneficial for usage during the startup phase of production when process parameters are the most difficult to track and interactions the least understood. Additionally, because some variable relationships are known, but not easily quantified, DSM's binary entry system is ideal for tracking qualitative and quantitative correlations together.

However, DSM has one significant pitfall which needs to be highlighted before it is used in start-up analysis. Because it is assumed that the interactions are known and stable[6], if the system is fluctuating there can be issues surrounding DSM's long-term validity. This weakness forces DSM to be revisited when information is upgraded or changed. Within the data management model given in Figure 1, a DSM mapping of the process variable interactions occurs every time the cycle begins the modeling portion of the procedure. DSM can then be considered a sort of "snapshot" of current correlation knowledge within the process – with the understanding that it can change especially in a start-up situation.

[6] Welch.

2.2 Approaches to Statistical Process Control

First used by W. A. Shewhart of Bell Telephone Laboratories[7] and later popularized by William Deming in Japan, SPC encompasses the display of data from established processes. The cornerstone of these methods is the ability to compare current data with known "good" data, that is, data from the process when it is producing within specifications. Then, with an appropriate basis for comparison, variation within the process is quantified.

Process variability can be characterized into two major categories: common cause variability and special cause variability[8]. Common cause variability encompasses variation present naturally within the process itself. This type of variation is random and cannot be controlled without changes to the process itself. Examples of this type of variation include sources ranging from different machine tolerances and capabilities as well as measurement error. Common cause variability is contrasted by special cause variability which results from some process change that is not normally present. If a difference in run conditions or raw materials is detected, these would be examples of special cause variability.

Typically, variability is determined through large amounts of data being analyzed through a company specific process, sometimes referred to as a Gage Repeatability and Reliability Test, or Gage R & R. Depending on the requirements at the given process step, data is generated using typical variables such as different raw material lots and shifts of operators. This data is carefully gathered in the hope that only common cause variability is present. With this data, many process data descriptors can be generated in order to better understand the natural process variability and the resulting capability.

2.2.1 Quantifying Process Variability

Given any set of data, certain attributes describing the variability within the data set are easily determined. The most prominent of these are the data mean and variance. The mean of data set j, μ_j is given by the equation below where there are N variables measuring variable x:

[7] Besterfield.

[8] Hawkins, et al.

$$Mean = \mu = \frac{1}{N}\sum_{j=1}^{N} x_j$$

Using the same notation, the sample variance is then:

$$s_j^2 = \sum_{j=1}^{N} \frac{(x_j - \mu)^2}{N-1}$$

As can be seen within the formulas themselves, the mean is an expression of the average value, or expected value, of the process variable while the sample variance give a measure of the deviation from this expected value for the given data set. Both of these calculations are subject to error and depending on the size of the data set, the true mean and variance could be significantly different than what is calculated. But, in the limit of an infinite amount of data, the variance is a measure of the true process variability, often expressed as σ^2. The term "standard deviation" is simply the square root of the variance and is often expressed as σ.

Another useful measure of the data set characteristics involves the median. The median is defined as the middle data point when the data is listed in ascending order. With an infinite data set and normally distributed data, the mean and median should be an equivalent measure of the expected value of the process variable. But in normal circumstances where data is not infinite, comparison of the median and the mean provides a measure of the skew of the data[9]. An additional feature of the median is that it is outlier resistant[10]. If a data set has a few data points far from the true expected value of the process, the mean will be greatly affected as each data point is equally weighted. The median, however, will not show much of a difference with the inclusion of these data points.

With some additional assumptions regarding the underlying distribution of the data, these basic descriptors for a given data set can be used in more sophisticated analysis of the process itself. If it is assumed that the underlying data distribution is random and follows the normal distribution, several more process characteristics can be determined. In order to assume that the underlying data results from a normal distribution, the central limit theorem is employed. The central limit

---

[9] Alwan.

[10] Sall, et.al.

theorem states that for a random variable, after enough samples are taken, the mean of the resulting normal distribution will approach the true mean[11].

One set of analysis stemming from the normal distribution assumption involves the standard deviation of the normal distribution. Using basic probability theory, it is known that a data point should fall within three standard deviations of the mean 99.7% of the time. If this is an acceptable goal for the process variable, control limits can be set using this as a guideline. Upper and lower control limits are typically defined as the limits where the process should show its natural variability within a given confidence. In the above example, if one wants 99.7% certainty that the resulting product is within common cause variability, the measurement would be within three standard deviations of the mean. The common industry term "Six Sigma" (pioneered by Motorola) refers to the ability of a process to output product falling within six sigma of each side of the mean while still remaining within specification.

Building on this, process capability indices can be calculated in order to help describe the underlying data distribution in relation to the product specifications. Process capability, or C_p, is calculated with the following formula:

$$C_p = \frac{UpperSpec - LowerSpec}{6\sigma}$$

As can be observed from the formula, this is a ratio of the width of the process variable specification to the expected width of the variable's distribution. If C_p is greater than or equal to unity then the process is capable of delivering within specification 99.7% of the time[12]. A more detailed measurement of the distribution characteristics involves looking at the upper and lower end of the specification with relation to the mean of the data set distribution. These measurements are commonly referred to as C_{pk} and are expressed by the following equations:

$$C_{pk,upper} = \frac{UpperSpec - \mu}{3\sigma} \qquad C_{pk,lower} = \frac{\mu - LowerSpec}{3\sigma}$$

$$C_{pk,overall} = \min(C_{pk,upper}, C_{pk,lower})$$

[11] Alwan.

[12] Wise, et. al.

If either of these is negative, it is quickly observed that the mean is outside of the respective specification. Also, these calculations reflect if the distribution is skewed to one side of the specifications more than the other. Process capability calculations provide a useful tool to evaluate both the process performance and the validity of the specifications.

2.2.2 Displaying Process Variability, Single Variables

Once a given process has been defined in terms of "normal" or "in-spec" conditions with the above criteria, current performance can be tracked with charts to help visualize any movements from the mean. The most basic set of charts are the X-bar and Range charts. The X-bar chart takes sets of measurements, usually in groups of five or ten, and plots them against the predetermined mean and control limits. The size of the data groups is based on the process itself and how the different data points can be related, and therefore similar enough to plot as one point. In cases where each process data point can be considered independent from the one preceding and following it, the X-bar chart becomes known as the individual response, or IR-chart. The Range chart is the plot of the difference of the current data point from the preceding data point. This graph helps to show movements in a single direction quickly, thereby triggering corrective action sooner.

Along with these charts, certain statistical "rules" are placed on the data in order to determine if the variability appears to be following random process fluctuations. Many different versions exist and the application of these rules can drastically affect the false alarm rate of the SPC tool. The false alarm rate is the number of times the SPC tool claims that the data is "out-of-control" due to special cause variability when the variability is actually random. This is also called the alpha risk, or Type I error, in some texts. Similarly, a beta risk (or Type II error) in the SPC analysis is when a bad lot of material is not flagged. Minimizing the beta risk lowers the risk of "bad" product reaching the customer while reducing the alpha risk increases the cost of production by reanalyzing "good" product[13].

An extension of the Range chart's capability, the Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts a display more sensitive to subtle process

[13] Griffith.

movements. The CUSUM chart plots the sum of the deviations from the mean for each subsequent measurement, as expressed by the following equation:

$$C_n = \sum_{j=1}^{N} (x_j - \mu)$$

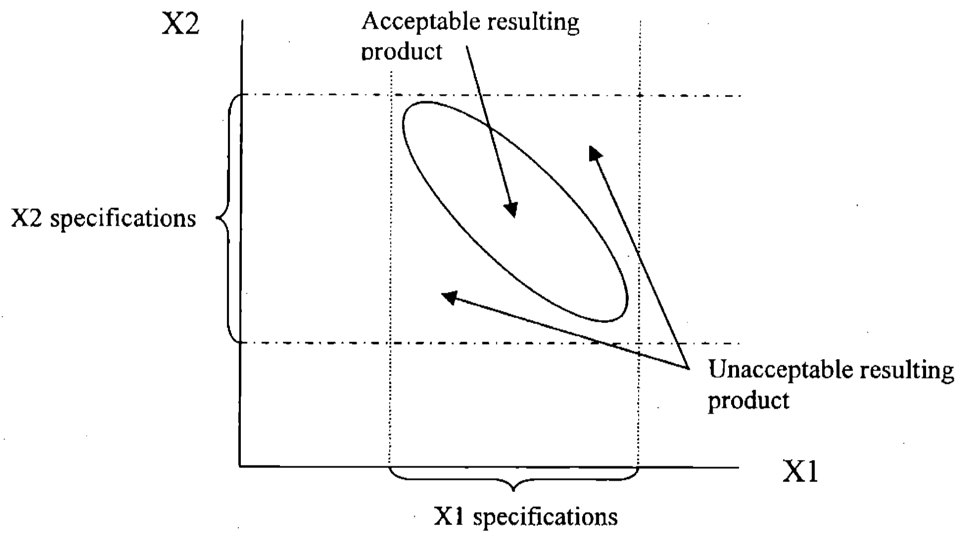
The EWMA weights the most current points with a multiplier in order accentuate any new process shifts. Both of these charts rely heavily on the assumptions that the data is normally distributed and the mean and standard deviation is known with a high degree of certainty. Because of the strong dependence on these assumptions these charts are generally unsuitable for the uncertain data sets present during start-up conditions, but work well with long-established processes in need of fast response to process changes.

2.3.3 Multivariate Process Monitoring

In some instances, a process variable that controls quality is the parameter alone. In these cases, it is the relationship of certain process variables to other parameters that determine the final outcome for the product. The following figure illustrates[14] an example where simple one dimensional specification in two variables does not account for an important interaction between the variables which ultimately affects the final quality of the product.

[14] Wurl, et. al.

Figure 5: Basic Illustration of Bivariate Relationship[15]



Many different SPC display types attempt to provide ways to control the interaction of variables, but one of the most straightforward is named the Projection to Latent Structures (PLS) method[16]. The PLS method uses the assumption that most of the process variables are highly correlated and can therefore be monitored by forming linear combinations of the individual process variables and parameters. Then, by using these grouped "super" process variables, traditional SPC methods can be applied.

As an example, take a collection of process condition variables and arrange them in a vector from X_1 to X_j. Similarly, form a Y vector (with k attributes) from the resulting product attributes of vector X. If there are m observations of data, a matrix of X vectors can be formed with the dimension of m x j as well as a Y matrix with dimension m x k. Using the PLS method, these matrices are then reduced to independent combinations of the original X and Y vectors, yielding the final PLS process condition variables and resulting product attributes. Depending on the initial number of process variables and final product attributes, the reduction in variables to monitor during the manufacturing cycle can be greatly reduced. The key drawback is that the resulting variables may be difficult to understand as they are not directly tied to a physical attribute.

[15] Wurl et.al.
[16] Wurl et.al.

Before using this method, it must be acknowledged that the PLS process condition variables and resulting product attributes contain some imprecision as very few process variables are perfectly correlated with each other. Additionally, one of the most important requirement of PLS analysis is that the process condition data (such as machine speed or temperature) be matched exactly with the resulting product attribute data. The logistics of matching the data correctly can be logistically challenging. Taking these negatives into account, PLS still provides the valuable advantage of allowing traditional SPC charting techniques to be used.

This method provides a valuable glimpse into possible critical relationships early in the production cycle. Because start-up data is typically more susceptible to larger fluctuations, correlations between process variables can be established with a larger range of data. Additionally, because of the simplifying nature of the PLS method, SPC can be initiated earlier in the production cycle, thereby creating a means to monitor process behavior. And, because much of the start-up data is discarded in the first place, this method's imprecision does not necessarily detract from the quality of the underlying data.

2.3.4 Traditional Short-Run SPC Solutions

Because of SPC's large data requirement, use during shorter production runs or during the beginning of production cycles has been minimal. In fact, Shewart himself proclaimed, "Control of this kind cannot be reached in a day. It can not be reached in the production of a product in which only a few pieces are manufactured. It can, however, be approached scientifically in a continuing mass production."[17] However, some degree of process control can be attained during short runs or the initial phase of a higher volume manufacturing cycle.

1. CUSUM or EWMA charts provide more sensitive responses to the process thereby detecting change quickly.
2. Modification of the standard control limits can lower the increased false alarm rates related to analysis using smaller data sets.
3. Stricter monitoring of process condition variables instead of product attribute variables allows for more accurate control charting while providing a stable basis for product attribute variability to express itself.

[17] Alwan.

4. Aggregating data from several sources, be it different lots, or pre-production data, can allow for traditional Shewart charts to be plotted correctly.
5. Following more robust forms of control charting to adequately shield results from the effects of outliers.

Using any of these choices[18] separately or in combination can allow for control with fewer data points, thereby easing one of the restrictions of traditional control chart application to shorter runs or start-up manufacturing scenarios.

2.3.5 Summary

Displaying and monitoring are the foundations of process control, but the extent of each defines the type of process control installed. Process control can take many forms, as described in the book, Dynamic Manufacturing[19]. The first type, reactive control, is the most basic relying on final test results to determine the acceptability of the product. Next, preventative control extends reactive control by finding intermediate process variables that cause quality problems in the final product. These intermediate variables are monitored and tolerances are determined based on acceptable product runs. When the magnitudes as well as direction of the intermediate variable correlations are taken into account, the control scheme is defined as progressive. Finally, if factors other than final product quality are included in the control schemes, a process is under dynamic control. With each increasing level of process control, activity around the manufacturing process moves from purely reactive to highly proactive. This movement allows for higher quality as more of the critical factors are determined as well as more productive analysis time.

2.4 Theoretical Background Summary

In any manufacturing environment, whether it is start-up or normal run conditions, the one constant is that the output is desired to be within a certain set of specifications. Ideally, a robust process model exists by the time production first starts and all necessary intermediate process tolerances are known and the ways to control them are immediately apparent. This is very rarely the case. Instead, it is required more frequently that process models are developed while production begins and the controlling variables are found along the way. Using the fundamental

---

[18] Ibid.

[19] Hayes, et. al.

ideas described in the previous section, data from a start-up production cycle can be used to quickly develop process models adequate for initial manufacturing. Then, based on these approximations, control schemes can be generated for the time between start-up and full-scale production.

3.0 Data Collection and Analysis Strategies

The previous section provided an introduction to three powerful data analysis tools: Design Structure Matrix (DSM), Statistical Process Control (SPC), and Projection to Latent Structures (PLS). Combining DSM methodology with PLS allows individuals involved with start-up data analysis to quickly reduce the number of variables that require attention. Then, with the use of SPC, these critical process variables can be displayed and used for controlling the process better while a more precise process model and SPC strategy are developed.

No analysis can begin without adequate data. The following section begins with a guide to the type, quality and form of the data necessary to complete initial process data modeling and SPC. After the data and information systems are understood, a discussion of how to map the process and the information it contains follows. Finally, the three preceding methods are brought together in a framework for efficient process modeling and control during production start-up.

3.1 Data Collection Strategies

Collecting data in a useful manner is a large challenge in any production situation. Just as it is becoming easier to generate reports with thousands of measured parameters per part, it is more and more difficult to collect it in a manner meant for display and analysis. Ideally, systems would be set up before production starts, yielding instant feedback. While this is critical for quick reaction to process trends, it is difficult to plan ahead in most new product situations. However, if an effort is made before production, many critical failures can be avoided.

3.1.1 Team Structure

Similar to many other large projects, creating a system for data collection an analysis requires a focal point and project champion. It is helpful for a team leader to know all the steps to the process and to be comfortable with most aspects of the company's information technology. With even a shallow knowledge of both the production process and information technology, this individual will be able to coordinate the data collection effort efficiently.

Key to any large data capture is the information technology capability of the organization. A simple map of each individual's computer literacy skills help to allocate work effectively for the project. An example of this follows:

Table 1: Resource allocation table enabling project assignments

Name	Low level programming (Machine)	Mid level	Networking	Web
Joe	none	C, VB	user	expert
Jane	expert	C	user	none
Tom	non	C	none	none

Besides allocating work, this will provide an additional tool for cross training, as in a fluid job market some skills are hard to retain. Additionally, any gaps found can be budgeted for contract work or designed into the future system.

Understanding the needs for the data add another level of complexity to the project planning. Depending on the scope of the production process, there could be many "process owners" along the path to final production. Each process owner must be willing to share knowledge and access to their systems before the project can commence. In certain corporate situations, this can be a volatile matter. If data is not currently public within a company, many individuals fear release of their data, especially to their supervisors. Without process owner support, any data collection effort is paralyzed; therefore, it is imperative that involved management communicates that any data collection initiative is only to identify those areas that need help, not to identify areas that have problems.

3.1.2 Creating a Data Structure

The data structure design begins once the team is assembled. In order to effectively structure any data collection effort, several key goals must be established. The highest level of decision is the data's usage. Many options exist such as:

- Process improvement
- Quality improvement
- Efficiency improvement
- Process definition

Obviously, if the underlying process controls are not understood, process definition must exist before any of the other choices are attempted. This is true in more than just "new" processes. Many processes evolve over time and the parameters that used to be measures of control may shift. Predictability is the only test of true process definition. If the inputs are known, any good process model should predict the outcome within a predetermined tolerance.

In the case where process definition is the initial goal, the structure of data collection is critical, but easier to identify. Because no process parameter can be initially excluded, all available data should be gathered in a central location. For instance, in a process with several data generating fixtures, a flat database can be constructed. The benefits of this strategy are that no possible process variable is excluded subjectively and there is no question of where each piece of data goes. Unfortunately, the downside is that the volume of data can be cumbersome to control. Even past the amount of disk space required to retain the data, efforts around automating the flow of data from its measured location to the central storage location can be challenging.

In all other data collection strategy cases, the need for the data can be better prioritized and more efficiently implemented. If the core use of data is for efficiency improvement, focus might be around cycle time of each fixture. Similarly, if quality improvement is the goal, cause/effect relations become important. In the quality improvement case, subjective data is a challenge by itself.

Over time, the data use may change or evolve into something different. Because of this, flexibility of the data structure requires initial consideration. If the first task is to define the critical process parameters, a natural follow-on would be to improve the process for efficiency or quality. So, while a large, flat data structure might be the first step, a more dispersed and focused structure might be the final form of data. In all instances, the future needs for the data leads to the proper design. As long as this is agreed upon before the task is initiated, resources will be better utilized for the successful outcome of the project.

3.1.2.1 "Firefighting" Data Structure

Historically, much of the software used in manufacturing was incapable of interfacing with other software involved with the process. Because of this, data generated at each particular fixture tended to be location or process step specific. This data, because of its isolated nature was good for little more than problem investigation of large failures.

It is still very common to find process data in individual reports on local machines. These data reports typically come in a form specific to a period of time, a specific product number, or run number. Typically each report has standard header information with process data following. The file name usually has an indicator of the specifics associated with the report, such as a date/time stamp of serial number. This form of data is especially inefficient due to two factors. First, because the header information is normally identical between many of the reports, this redundant information tends to consume resources such as disk or screen space. Secondly, each result resides on a separate page which removes any information from the previous run. This, in effect, makes the history of the process fixture difficult to access while using the fixture on the next process iteration.

If the "firefighting" data structure is prevalent in a process area, very little proactive analysis can take place. Because the individual reports are normally not appended to each other, data for a particular process parameter might exist in hundreds of different files. Due to the large amount of effort required in collecting multiple reports together, it is normally not typical for this data to be used on a daily basis.

For the majority of cases, a data structure of individual reports on local process machines has very little use beyond inquiry for very particular problems. The only additional benefit from this type of data structure is the ability to remain an independent process fixture while using the pre-existing software or connections. In general, "firefighting" data structures make the installation of SPC extremely difficult and proactive data structures are preferred.

3.1.2.2 Proactive analysis structure

Ideally, data generated at particular manufacturing fixtures would be in a form that allows multiple types of analysis – both economic and technical. To attain such flexibility, the first requirement is for the data to be accessible. This immediately precludes the use of individual text reports on local machines as an option. Instead, as discussed before, data collected in a common area, such as a database networked to other computers is a method to create a useful window to the process information.

Allowing accessibility to the data is the first step to finding value in data. By encouraging different individuals from both within and outside the affected area to look at the data, powerful relationships might be uncovered. If process variables are considered as part of the entire system, it might have more meaning than at its own step. Working with data further upstream and downstream helps knowledge of the overall process.

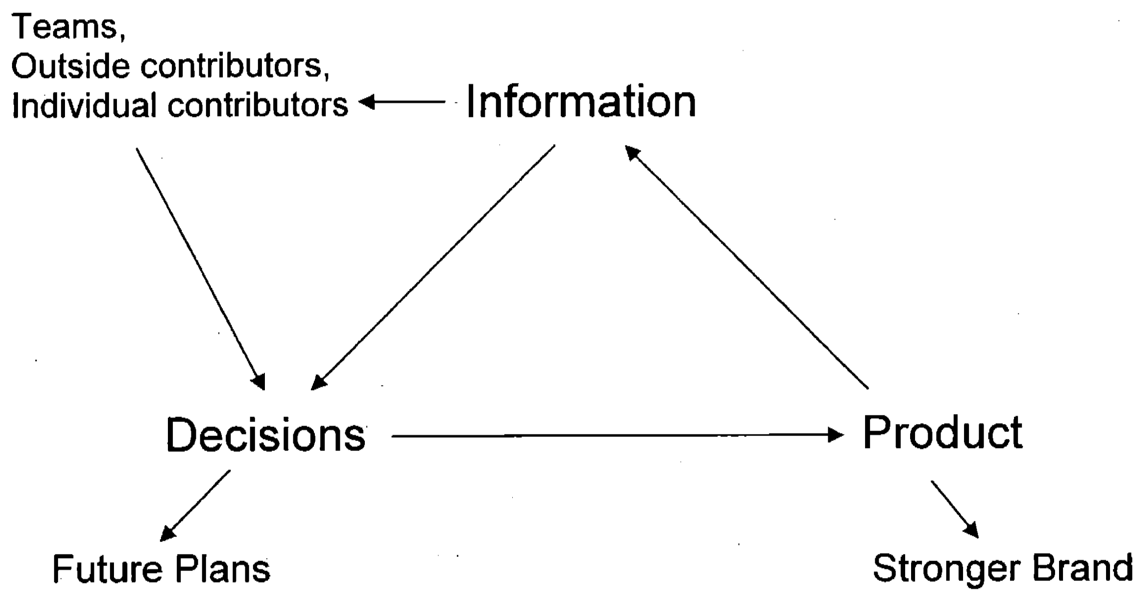
Data accessible to many people from one place is only the beginning. The next consideration is the form that data should take. If process modeling is the desired outcome, various parameters for the same unit should be collected together. This facilitates analysis of interactions between variables on the same unit and trends over time. Then, in turn, displaying the data through SPC allows individuals to easily look for relationships and trends that could benefit the overall process.

3.1.3 Summary

Building access to data that can be analyzed proactively is the first step to creating a self-reinforcing loop centered on making better decisions. Once the data is available for team members to observe and analyze, more and better input can be used in decisions affecting everything from everyday run conditions to the next product development cycle. Then, the resulting alterations from the decision generate more data for analysis. This cycle builds on itself creating value through efficient decision-making, higher quality product, and improved future designs.

Information formed into a proactive structure as discussed earlier facilitates this cycle. Unlike individual unit reports that must be compiled before decisions, data arranged in an easily accessible form encourages a proactive analysis behavior. In some instances, arranging a system where employees can easily access information facilitates a shift in culture. Regardless of the company culture, more informed input into daily decisions increases incremental improvement as well as longer term projects. As long as the channels between the information, decision making and product design and manufacturing are open this cycle is self-reinforcing.

Figure 6: Self-reinforcing decision making cycle



3.2 Overall Process Data Mapping Approaches

3.2.1 Background

Modeling a production process is complicated, yet critical to efficiently making high quality product. Without adequate process modeling, capital expenses are not only more expensive, but can also be ineffective at producing the desired product. Additionally, a well understood process can be controlled to maximize the output's value to the market. And most importantly, a defined process has limits that are enforced for the increased safety of the employees working with the equipment. For all of these factors and more, defining a full process model is the first step to making a successful product.

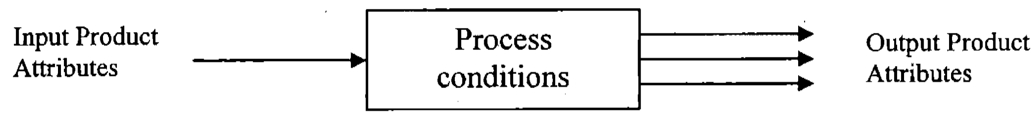
The level of model complexity grows with product complexity. So, as products become more technically advanced, the effort needed to fully define the process increases. Luckily, because most products have similar underpinnings to previous versions, older process models provide a valuable start. But, when a product/process is sufficiently different that beginning from an older model is too cumbersome, a new model is required. This can be a daunting task as with the advent of computers, more data is available today than ever before. More data is not necessarily better data, however, and determining the critical process parameters can be muddled with extraneous information.

As product cycles shorten to meet differentiated customer demands, there is less time to analyze new iterations. This can lead to misleading thought models based on older products which may not be applicable for the current product. Decisions based on these outdated models may not be sufficient for profitable operation. Instead, a systematic and thorough way to approach each new product is necessary to ensure that current requirements are met efficiently.

To start a large process model, the first step is to understand the information that is available and approach it in a consistent and thorough manner. As described in the Data Structure section, not all process and product data exists in an easily accessible state. Once the form of data is understood, it must be matched at each process step to ensure that any correlations that develop are between the appropriate variables. The most common way to accomplish this is by matching the numerous variables by each product unit through the system.

A necessary concurrent effort while gathering information on the form of the process data is to categorize the variable types. The most basic categorization is to group the variables by whether they are measured attributes of the product or process conditions. Product attributes encompass a wide variety of possibilities from incoming raw material measurements to final product performance results. Process conditions can be defined as measurements of an action done to the product that might change, remove or create a product attribute. The following diagram illustrates a model process step where certain product attributes are exposed to measured process conditions which result in a set of output product attributes.

Figure 7: Diagram showing the relationship of process conditions and product attributes at a process step.



When approaching the beginning of a process, the raw materials present a critical set of input variables. Some areas to research during the initial project phase are the following:

- Number of qualified suppliers
- Incoming specification ranges
- Batch-to-batch variability
- Incoming quality audits vs. vendor measurements

All of these considerations can help to define the incoming variability and therefore, the robustness of the process under consideration. Collecting a list of qualified producers will encourage more questions as most producers have different specifications and process capabilities. Understanding the repercussions of all the incoming specification ranges can prove important when running sensitivity tests on the final process model. Similarly, batch-to-batch variation affects production, especially if a process is sensitive to one or more material attributes. Finally, ensuring that the vendor's measurements are consistent with the new process' measuring capability is analogous to people communicating with the same language. If the new process is assuming a specification range based on a different test method or with a method which incurs an offset, this should be considered within the model.

More information normally exists around the end of a process. Because the new product has been defined using customer needs and perhaps previous products, requirements are more lucid at this point. The only exception is when customer requirements are qualitative in nature. In this case, great care needs to be taken to ensure that qualitative expectations are correlated to quantitative process information. This is the only consistent way to monitor a process.

Once the start and finish of a process is defined, the task remains to link the two. In most cases, this requires analysis of many intermediate steps. Also, during this process, one may find that parameters that must be correlated along each process step require additional data somewhere along the chain of events. Thus, process modeling includes information not readily available as well as accessible data. Many times it can be a trivial task to obtain further data. For instance, in many cases, suppliers of raw material have manufacturing specifications beyond sales specifications. If the process modeling requires more information at the first step, simply requesting additional product information can aid in the modeling effort. Additionally, because process monitoring equipment is more advanced, many parameters may be measured but not displayed or captured. Understanding the capability of the processing equipment is an additional important preparation for process modeling.

Even with thorough efforts around the process end points, further prioritization may be required. Many times, it becomes necessary to temporarily focus efforts on certain trouble areas or build support for the modeling effort by showing immediate results. Two obvious areas to begin with are the process bottleneck and the customer critical specifications. Because knowledge around the threshold of the process bottleneck yields information about increasing throughput, it can be a quick source of valuable results. Linking the bottleneck process performance to customer specifications is the major goal if increasing throughput is the desired outcome. Many times, customer specifications are determined during process steps other than the bottleneck. If this is the case, understanding the performance of these customer defining steps increase quality and therefore customer satisfaction.

3.2.2 Single Stage Analysis

Before an entire process model can be established, each stage must be independently defined. This thorough approach begins by compiling a list of all measured attributes and, if available, the general final specifications the attributes affect. If the later does not exist, it is of little concern as the process model will develop the important relationships once completed. Once the list of step attributes is collected, each one must be analyzed separately.

First, a simple time series plot of each attribute will reveal how consistent the parameter runs over time. Then, simple statistics will calculate the mean and variation of the data set. Understanding the variation in relation to the mean is a useful tool for control. Certain process parameters will be more sensitive to change and therefore provide effective tuning mechanisms for Digital Control Systems (DCS) or Statistical Process Control (SPC). Conversely, a process parameter that appears to remain very stable might provide a valuable insight when excursions are experienced. Identifying the outliers and how they relate to special causes provides valuable information. First, identifying any negative or positive outcomes from the change will allow prioritization of response. If nothing resulted from the change, perhaps specifications can be widened. However, if drastic manifestations of the cause result, more direct attention can be given for either a change in process or product design.

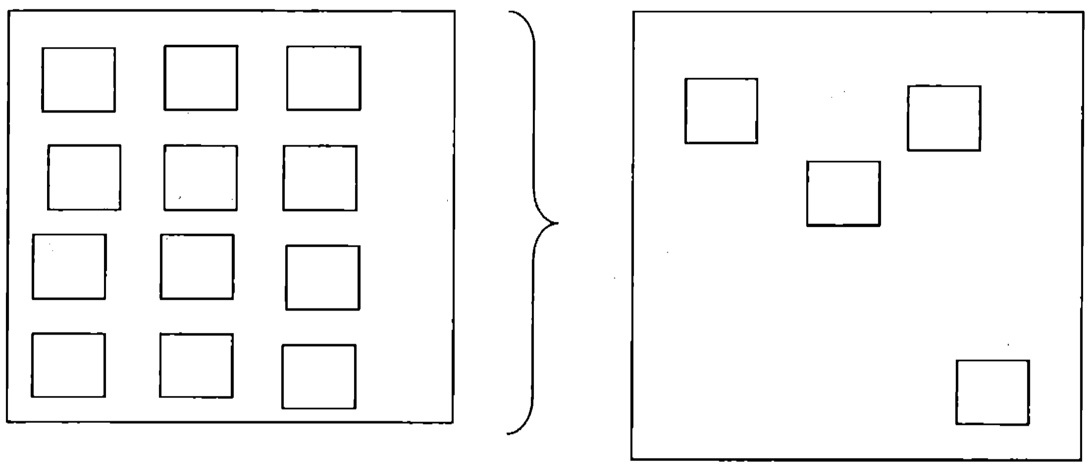
Finally, each individual variable must have a measurement confidence level within an acceptable range. By completing a design of experiments using criteria such as multiple technicians running tests, test fixtures, and lots of raw materials will quantify the capability of each measurement. Without this effort, the process might be characterized with noise instead of true signals, making it difficult to control.

Once each individual variable is characterized, they can be related to one another. This step reduces redundancy, thereby making everything from control to databases easier to manage. Using any basic statistics package, the process parameter should be sequentially plotted against each other to identify any strong correlations. Some process attributes can determine several others, these are Process Step Critical Variables (PSCV). The criteria that PSCVs need to meet are the following:

1. Physically explain one or more other process parameters at the same step
2. Correlate with one or more process variables by a model with acceptable error, even without physical explanation
3. Independent of any other process variable being considered at the time

For a process parameter to be considered a PSCV, it must satisfy any of 1) and 2) or 3). The first two requirements are fairly obvious, but the later is more subtle. If a variable appears to be independent within the process step, it must be considered for further analysis. Many times, those variables that do not have any obvious correlations have effects further downstream in the process. Using these criteria will allow the process analyst to select the parameters with the most promise while discarding redundant measurements. Also, if a variable has outliers which correlate to drastic events (whether characterized by safety or product quality), it may receive a higher prioritization than another variable that it correlates to in order to capture this failure state.

Figure 8: By eliminating redundancy within a single process stage, variables can be selected for further analysis



During any type of statistical analysis, there is uncertainty. While defining PSCVs can reduce redundancy, the process itself can introduce more uncertainty as process parameters are removed from consideration. Creating rules for accepting correlations can mitigate the negative effects of deselecting process parameters. These rules allow objective evaluation of any resulting process model which can simplify the modeling process. When comparing process parameters at the same process step, the goal is to remove redundancy but maintain any independent variables for further analysis with other process steps. Therefore, this stage of the analysis should have strict guidelines for removing a variable from further consideration. In the best cases, correlations should be very strong to ensure that no valuable information is lost. Using standard statistical measures of correlation strength such as R-squared values allows for objective assessment of a model's fit.

The analysis process itself is generally straightforward. With matching data for all of the process parameters at the chosen stage, one variable should be singled out each time and simultaneously compared with all the other variables. Depending on the complexity of the stage, the data can be broken into subsets based on test results or by result type. For instance, if there are five separate tests run at a stage, each generating twenty variables, it might make sense to analyze the five separate test results first and then look at the overall stage once some variables have been taken out of consideration.

Misleading results can be generated when modeling a single response from all the other step variables. Because statistical modeling software is so sophisticated, many packages can generate models with apparently high correlation when none truly exists. This is a danger with working with too many variables at once. In general, if a model looks too complicated, it probably is too complicated. Before attempting a process model, it is worthwhile to research the most complex model in a related process to set a guideline on how many process variables can realistically be explanatory variables.

If two process parameters appear to correlate well with one another, there can be some difficulty choosing which to discard and which to keep as a PSCV. Before either is set aside, the first hurdle to clear is explanation of outliers. Outliers in this sense can provide valuable insight into critical relationships required for an acceptable product. In the case where severe outliers create unacceptable product, both variables need to be tracked further to ensure high quality product. Bivariate SPC charts might become a valuable tool in many of these cases.

Additionally, interactions can be another complicating factor. In many cases, there is little physical explanation for most interactions. Because many are hard to comprehend, it may be important to check for interactions where none are expected. And, if interactions on a first order basis are found, higher order interactions can be investigated to create a more thorough model. Again, including interactions may make models more complicated and there must be a guideline before beginning to ensure that the model is correct and not just a figment of the particular data set being analyzed.

To fully define each stage of a process requires painstaking attention to each process variable and its relationships to other variables. First, each process variable must be analyzed on its own. By looking at the basic statistics, one can determine how sensitive the variable is to process changes and evaluate the parameter's merits as a control variable. An analyst can also find valuable process information by finding the causes of outliers. Finally, by understanding the measurement error associated with each process parameter, the value of the measurement can be better quantified.

After each variable is studied on its own, then the stage can be analyzed as a collection of process parameters. By systematically attempting to find models for each of the process parameters based on other parameters within the same stage, a large amount of redundancy can be removed thereby simplify further modeling efforts. Before removing any process variables, guidelines for declaring a variable redundant must be defined and documented. Independent variables and variables with large negative consequences should automatically be classified as PSCVs and allowed on for further overall process analysis. Other considerations include acceptable error in correlations and significance of outliers. After each stage has undergone a thorough analysis, the PSCVs are used in the next phase of modeling, sequential stage analysis.

3.3.3 Sequential Stage Analysis

The next step in process modeling is relating sequential stages of the process. As the PSCVs are mapped at each stage, the next iteration is to relate PSCVs of different stages to each other. The goal of this exercise is to eventually track PSCVs upstream to their origin and downstream to their final value. Once the full analysis is complete, most sources of variability will be quantified enough to be controlled adequately by either Statistical Process Control or other control schemes.

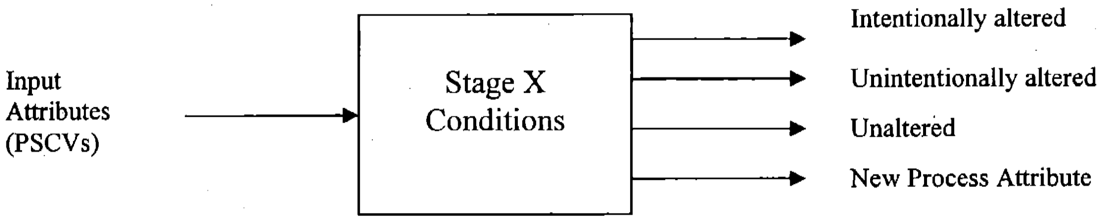
Sequential stages can be a misleading identifier, as many times different types of process attributes will be manipulated at stations separated from each other by other process steps. To prevent confusion, organizing process parameters by final specification category will help to categorize the various variables. Then, for each type of attribute, each process stage will fall into one of four types:

1. Benign process step
2. Intentional alteration process step
3. Unintentional alteration process step
4. Creation process step

Each process step will most likely fall into each of these four categories, but for different process variables. In most processes, different stages are designed to affect or create particular product attributes. So, an individual stage will be thought to intentionally affect certain parameters while being benign to others. Unfortunately, this is an ideal case and many product attributes are changed without knowledge. Using an assumption that process steps alter or create specifically what they are designed to can lead to many errors.

A stage diagram will help track all types of process changes. Each stage diagram can be considered a black box with inputs and outputs. Not only will this technique capture valuable information for planning, but it will also prove invaluable for process modeling, regardless of the software package. In the stage diagram, the labels descried above are observed from the stage point of view instead of the attribute vantage point.

Figure 9: A stage diagram tracks the attributes that are incoming and outgoing from each stage.



Certain parameters between stages will also show redundancy with other process variables. In many cases, new process attributes measured at a stage will be a function of the input PSCVs and the stage conditions. Additionally, those process variables altered unintentionally might be affected in a direction matching one of the other process variables. If this is the case, unintentional changes can be mapped by tracking the intentional changes only. By focusing on the parameters designed to change while understanding the effects on the other parameters, process understanding and control are more effective.

Quantifying models based on stage alterations is one of the first major steps to understanding the overall process performance. Whereas single stage analysis simply looks to reduce redundancy, sequential stage analysis strives to correlate parameters based on process changes.

3.3.4 Overall Process Descriptions

Once critical sequential steps are mapped according to PSCVs and their correlations with other process variables, an overall process description can be modeled. This is accomplished by bringing the stage diagrams together and fully mapping each variable throughout the process. Depending on the complexity of the system, this can be a large task. As mentioned previously, at this point in the analysis it might be wise to focus on a few critical attributes instead of all at once. By choosing parameters important for the end product or those which affect the operability of the bottleneck step, an analyst can show results in a more efficient manner.

While the same techniques as used in the single stage and sequential analysis can be broadened to the entire system, many times this is too cumbersome for efficient modeling. Instead, focusing on just the few variables that define the final product might prove to be a better approach. By the time sequential stage analysis is complete, the alterations made at each step are correlated with process conditions and many other minor process parameters. Using these as tools to model the critical few variables throughout the process shortens the modeling cycle.

Another approach to possibly lowering the number of process variables needed to describe the system is to look for variable relationships outside of each process step. Some of the more promising variable groupings follow:

1) Between the final product attributes
2) Between the final and intermediate product attributes, if stages not sequential
3) Between intermediate process step conditions, if stages not sequential

By trying these different variable groupings for analysis, additional redundancies can be found before a complete process model effort is undertaken.

If detailed analysis is desired, continuing on the path of sequential stage analysis might uncover more subtle process relations. Developing a systematic plan for correlation searches among all the process parameters at all the steps takes discipline and dedication. Linear relationships are easy to understand, but not common enough to base modeling efforts around. Instead, interactions between variables and nonlinear relationships need to be included to fully understand the overall process structure.

3.3.7 Combination of DSM and PLS methodologies

While narrowing the list of PSCVs, a consistent and organized procedure is required. By systematically going through the mapped process and identifying key relationships and variable redundancies, the time it takes to approximate the outcome of a process through a model is reduced. PLS provides a basis for tracking process variables that encompass many individual process variables. DSM will aid in this effort by displaying the results in a quickly digestible form that can be used in conjunction with an automated computer analysis program.

Analysis begins at the single stage analysis level. This level, as described earlier, involves the process test results and run conditions of a single process step. To begin the analysis effort, some standards specific to the process need to be defined. First, the units and statistical test used to measure statistical significance need to be determined. Depending on the conventions within the affected organization, this requirement can change. Secondly, what level of error can be tolerated using these tests? This is a balance between the desire to track fewer process variables and the need for relative predictive power. Considerations may include the cost of machine maintenance, personnel safety, or demand for product quality.

Once the initial conventions are decided, the analysis can begin. Using DSM as a framework for organization, the statistical measures of model fit can be arranged in a matrix of parameters. As an example, the following matrix contains the binary representation of the acceptable correlations which exist for nine process parameters:

Figure 10: Example matrix representation of model and observed process variables

		Observations									
		1	2	3	4	5	6	7	8	9	row sum
	1										0
	2			1		1			1		3
M	3	1				1				1	3
o	4						1	1			2
d	5		1	1					1		3
e	6				1					1	2
l	7				1				1	1	3
	8		1		1	1		1			4
	9	1		1			1	1			4
column sum		2	2	3	3	3	2	3	3	3	

Each row (labeled by "Model") represents the modeled process variable using the other process variables as explanatory variables. Each column contains the number of appearances the observations of that particular process variable in the respective model variable's mathematical representation. The sum of each row is the number of explanatory variables used for the model of that one particular process parameter. The sum of each column is equivalent to the number of times the observed variable is used in a model.

There are many observations about the data that become obvious with this representation of statistical relationships. First, if a row sum is zero, as in the case of model variable 1, then that process variable has no explanatory variables within the data set chosen for analysis. This type of variable can be termed independent from the variables contained within the group and should therefore advance to the next round of analysis. The next observation is that model variables that contain higher numbers of explanatory variables may warrant additional analysis. Typically, the higher the number of explanatory variables necessary for an acceptable fit, the less reliable the model is outside of the process condition ranges used for analysis. Also, in many cases, there will be a column sum that dwarfs many of the other observed process variables. This is indicative of variable redundancy where many of the process variables are not independent of each other.

In traditional DSM, this is the point where the matrix is rearranged in order to find the groups of variables with the strongest interactions. This exercise can be undertaken with process variables as well, although, since the matrix marks quantitative predictive models, the rearrangement effort is less necessary. Further, because the sum of the row and column elements are taken into account, rearrangement of entries onto the diagonal adds little to this analysis past a visual representation of the variables with the strongest interactions.

One additional benefit of this analysis is that with each process variable model, different outlier points will become obvious. While many of these are simply random in nature, occasionally these outliers will represent a consistent product failure. If this is the case, the model which shows the product failure as an outlier should become a candidate for special SPC monitoring or process alerts. A compilation of all product failures without known causes should be accessible for this type of periphery analysis.

The next step in this modified DSM analysis is to reduce the number of process variables needed to describe the system. Using the matrix, certain process variables are chosen to move onto the next level of analysis (sequential-stage). The first candidates are the variables that are independent, like variable 1 in our example. If no independent variables exist, then the model variables that require the most explanatory variables make the next best candidates. When a variable is chosen, its "model" row is removed, as it requires no explanatory variables. Similarly, it's "Observation" column is also removed as it is given that it will be provided as an explanatory variable. Removing the independent process variable (1) and the two variables which require four explanatory variables (8, 9), the matrix is reduced to the following:

Figure 11: Results of one iteration of process variable reduction

	Observations						
Model		2	3	4	5	6	7	row sum
	2		1		1			2
	3				1			1
	4					1	1	2
	5	1	1					2
	6			1				1
	7			1				1
column sum		1	2	2	2	1	1	

After one iteration of selecting process variables to move onto the next stage of analysis, strong candidates are less clear. However, taking the process variables with the highest combined column and row sum (4, 5), the matrix is further reduced to the following:

Figure 12: Results of two iterations of process variable reduction

	Observations					
Model		2	3	6	7	row sum
	2		1			1
	3					0
	6					0
	7					0
column sum		0	1	0	0	

This smaller matrix now shows that the reduction process is at its last iteration. At this point, either process variable 2 or 3 may be selected with slight preference given to 3 as it had the highest initial combined sum of row and column entries. However, if there was a practical reason for selecting variable 2, it would be acceptable as well. Examples of this might include historical use of process variable 2 or an easier accessibility to data of the variable 2 type.

In this small exercise, the number of process variables necessary for analysis has been reduced from nine to three. As the next stage of statistical modeling begins, this will greatly aid in the speed of results as compared with including all variables in the analysis.

3.3.8 Residual Pattern Analysis

The term residuals refer to the difference between the measured data of a system and the predicted value from a model. Because any least squares analysis is dependent on assumptions about the error present in the system, the residuals must have particular characteristics in order for the model to be acceptable. First, the expected value of the residual error should be zero. Next, the residual errors should have a constant variance over the full range of measurements and model. Finally, the residual errors should be independent from each other and the process parameters being used in the model[20]. The result of these requirements is that no observable patterns should be discernable from a plot of the residuals and any process parameter or variable used in the analysis. Additionally, these residual values should form a normal distribution given their assumed random nature.

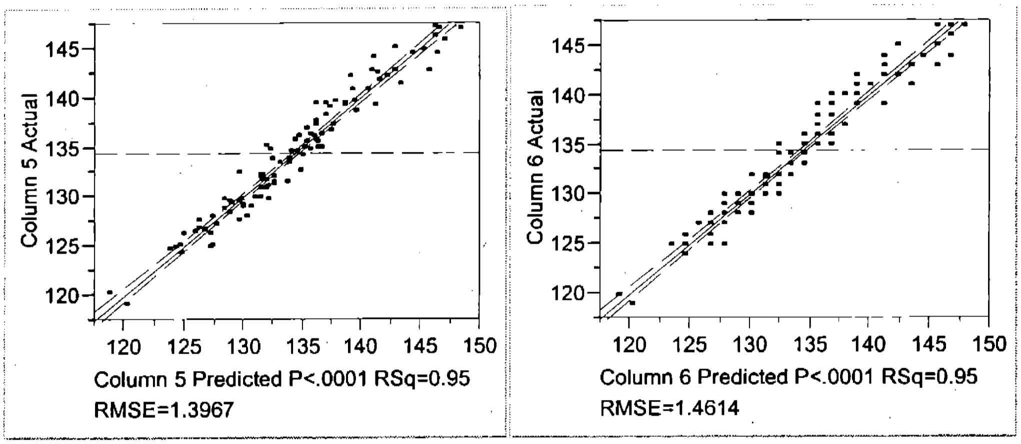
There are many common types of undesired residual patterns. For instance, many times the increasing magnitude of a process variable affects the error in the model. This type of residual pattern appears as arrow head or funnel when graphically plotted[21] and violates the requirement of constant variance. Similarly, in many cases the residual values drift over time. This is fairly difficult to diagnose if the data structure provided does not allow analysis chronologically. However, if the residuals are found to vary with time, this is a symptom of a missing explanatory variable within the model. Such cases necessitate further investigation into the missing factors needed to fully describe the process through a quantitative model.

Yet another pattern to look for appears in cases where at least one of the data sets is rounded. Examples are abundant – any time the precision on a gauge is less than another variable it is being related to, it can be considered "rounded" in comparison. For instance, many temperature gauges are in integer increments whereas a related pressure might have two are more places behind the decimal. Models based on these types of data sets can produce trends that appear to be striated. They look to be almost of a "lower resolution" as compared with their non-rounded counterparts.

[20] Vining.

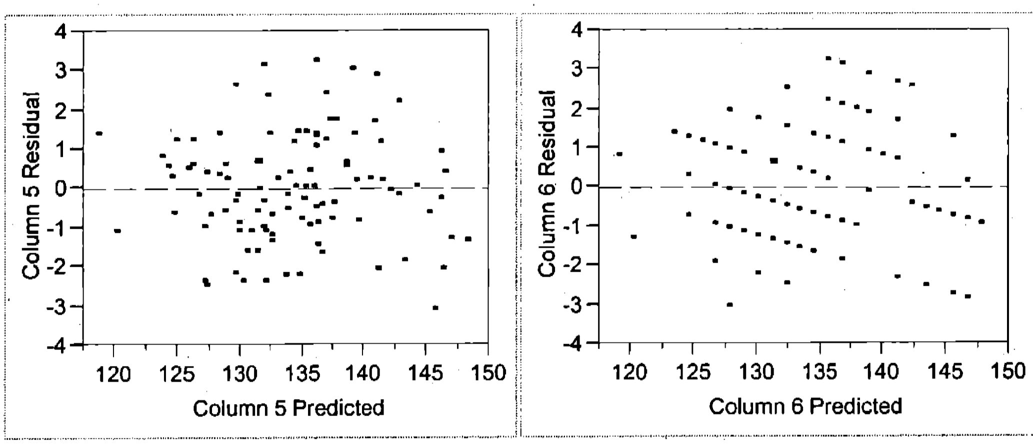
[21] Ibid.

Figure 13: Original data, with 2 places behind the decimal points and data rounded to the nearest whole number



So far, comparison between the previous two graphs is fairly obvious. However, the residual patterns give a slightly more intriguing view of the underlying model. The "lower resolution" data appears to have a consistent error in the slope, as evidenced by the slopes in the residual plot.

Figure 14: Original Data's residuals look random; Rounded data's residual appears to have a consistent error in the slope



To recreate this effect in a more controlled data set, the following example provides further insight into these "lower resolution" data sets. To start, one can establish a true underlying model. In this case, we start with a slope of unity and a decimal intercept.

$$y = x + 0.4 \qquad \text{(Equation 1)}$$

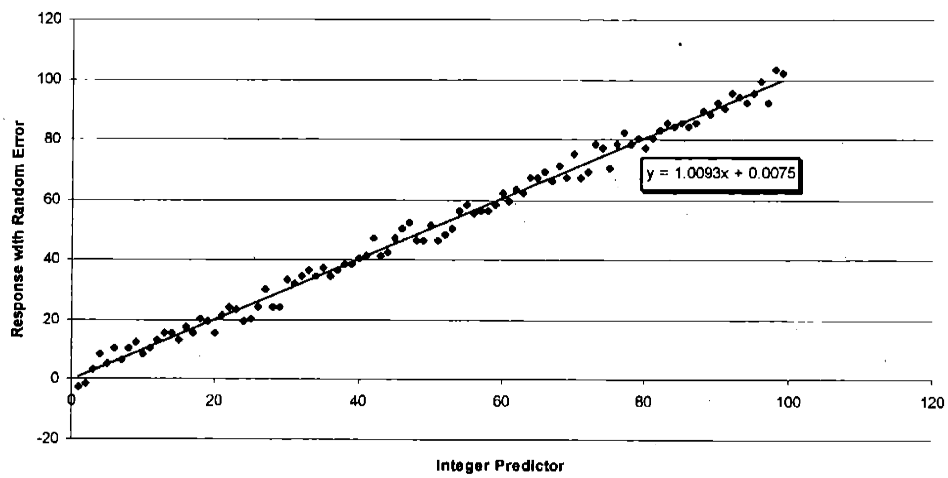
Then, using integers as the signal (x) and forcing random noise around the response, a data set is established with a known underlying model. Using integers approximates the "lower resolution" data sets encountered in practice, while forced random error is used to show natural variation.

$$y = x + 0.4 + \varepsilon \qquad \text{(Equation 2)}$$

After one iteration of random error, one can plot the response against the integer predictor. Then, using a typical statistics program, a model can be estimated based on the least squares method. This analysis yields a new equation, very close to the true underlying model.

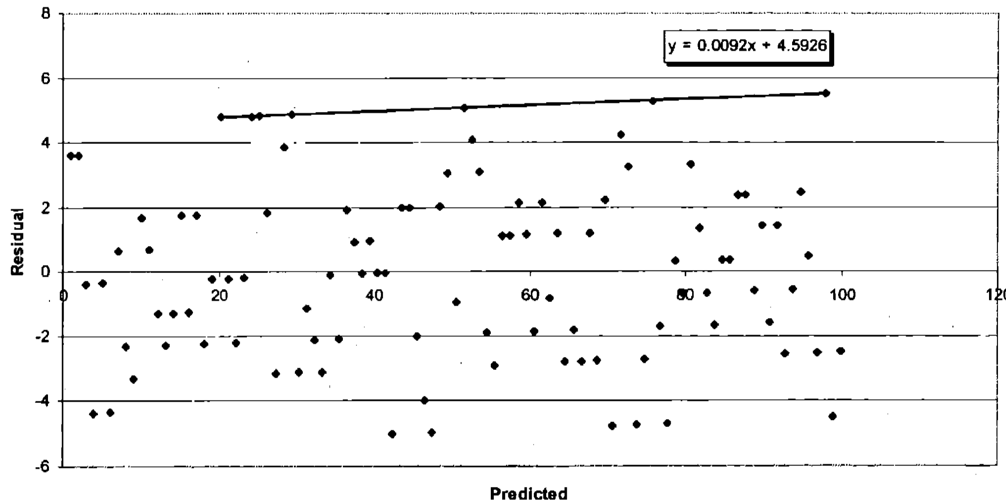
$$y = 1.0093x + 0.0075 \qquad \text{(Equation 3)}$$

Figure 15: Random error around Equation 1, Least squares fit of data



While the model produced using least squares gives an adequate fit, plotting the residuals again produces evidence of a systematic slope error. The residuals are plotted by subtracting the actual response, response of Equation 1 with random error included, from the predicted values. The predicted values are integers from 1 to 100 inserted into the model generated by least squares analysis.

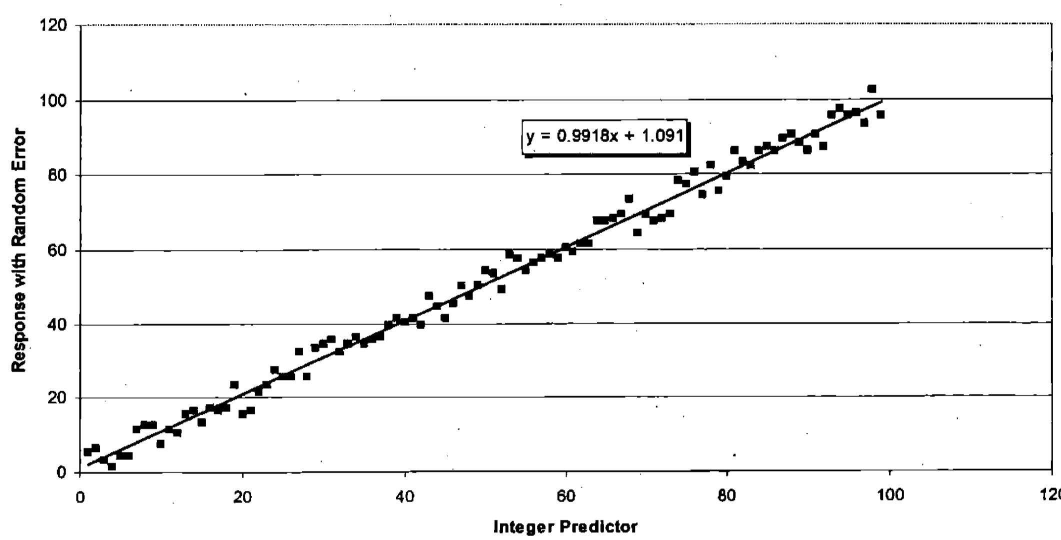
Figure 16: Slopes in the residuals indicate an error in the model



The systematic error shows an upward trend translating to a slope over-correction. As the predicted value increases, the residuals also increase. This is the expected response as the slope is known to be too high as compared with the underlying model. Because of the multiplicative error and the separation due to rounding, this residuals pattern acts as a magnification of slope error.

This also works in cases of models with slopes less than the true slope. Taking the same underlying model as in Equation 1, and introducing different random error until least squares fit produces a slope of less than one, one can produce a graph such as the following.

Figure 17: Random error around Equation 1, Least squares fit of data

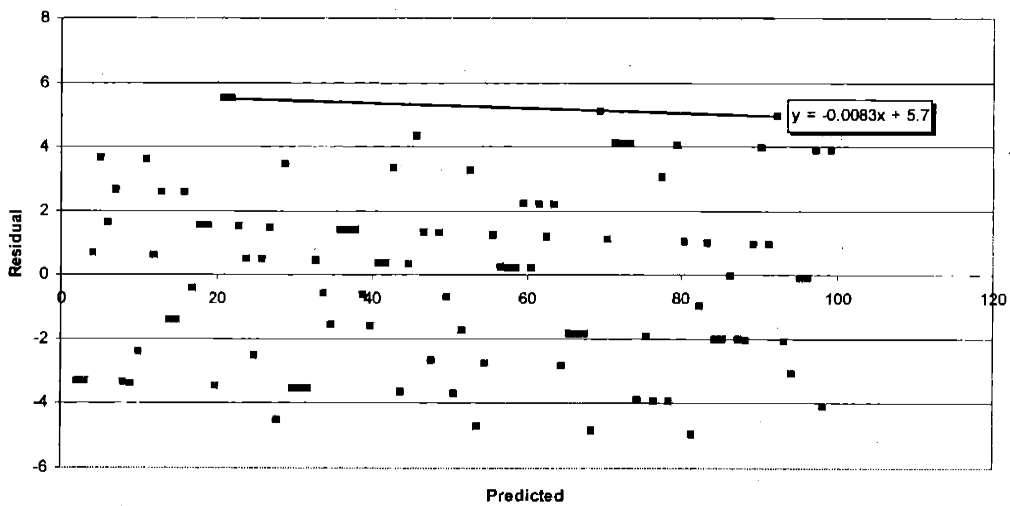


The new model for the data is expressed by Equation 4:

$$y = 0.9918x + 1.091 \qquad \text{(Equation 4)}$$

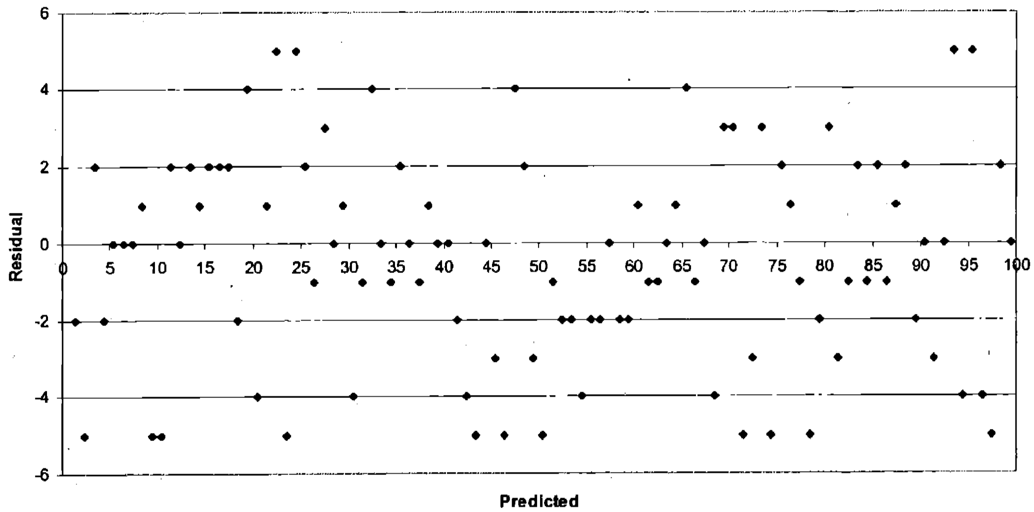
Again, this is very close to the model with a slope of unity, but using the residuals as a key to the slope error, we see the opposite effect of Equation 3's residuals. Our new model of Equation 4 produces a negative error in the residuals. Again, this is consistent with the idea that the residuals will become more negative as the slope is underestimating the underlying model (Equation 1).

Figure 18: Slopes in the residuals indicate an error in the model



If the least squares analysis calculated the slope exactly at unity and the residuals were plotted, the slope pattern is removed. When it is simply random error added to the underlying model, as expressed by Equation 2, the error is plotted with the same rounding separation, but without a systematic slope.

Figure 19: This graph exhibits a residual plot where the slope of the model equals the underlying model



If the data is rounded out of necessity or as an error seeking technique, observing the residuals provides a method of observing error in the estimated slope as expressed by ε_1 in Equation 5.

$$y = (m + \varepsilon_1)x + b + \varepsilon_2 \qquad \text{(Equation 5)}$$

This residual pattern, while not always observed, can be indicative of a systematic error in the predictive model's slope.

3.5 Control Charting for Production Start-up

3.5.1 Introduction

Control charts, based on statistical analysis first outlined in the 1930's, have become prevalent in most production settings. Because of their successful use in Japan during the 1950's, American companies quickly learned to apply this strategy to their products. However, production has changed significantly over the last 50 years. Whereas, in the 1950s large production lots of the same products were being tracked, the short product lifecycles and product version proliferation prevalent today complicate the traditional approach to SPC.

Another complicating factor is the amount of data available. Previously, data collection was relatively difficult. In order to justify the actual collection of an attribute, an engineer had to be sure that it was pertinent to the process itself. Today, the case is more of "data overload". Most processing equipment outputs more data than can be understood which confuses the process modeling effort. While this is a seemingly better situation than in the past, engineers must take care that the control charts they generate are meaningful and do not act as a source of frustration to those who must track the process.

The most widely accepted form of control charting is a single attribute chart displaying subgroup means and ranges against upper and lower control limits determined with large amounts of historical data. While this is applicable in environments with a controlled process that has existed for some time, in today's short product lifecycles, it is more and more difficult to accurately portray processes in this manner. Other control charts, such as the previously mentioned CUSUM and EWMA charts can be tailored to a high degree of sensitivity for a known process. Unfortunately, the higher degree of sensitivity can increase the false alarm incidents which exacerbate the already hectic start-up period. This leads to the conclusion that, in order to have some way to track a new process, a new form of control charting is desired.

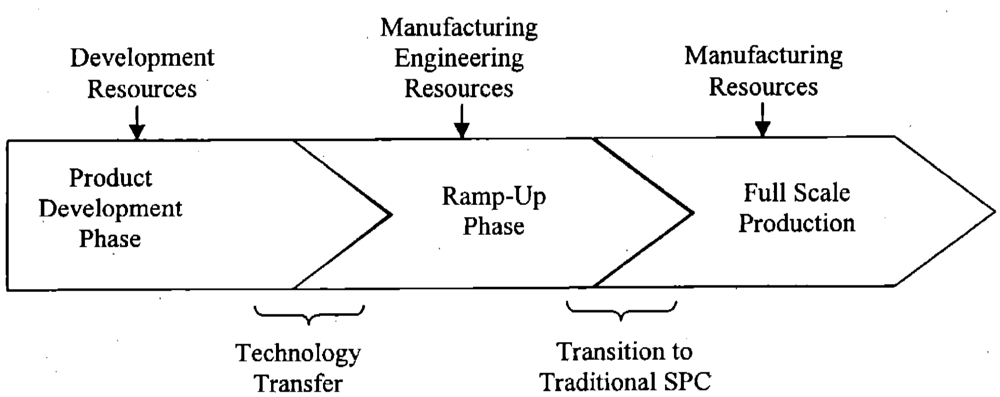
During production start-up, there is a tendency to participate in "fire-fighting" activities over careful analytical study of the process. This trend will likely continue as productivity expectations continue to increase and activities seen as lower priority are sacrificed in favor of immediate challenges. Because of this, the creation of a SPC-like display before the full process analysis is completed would be helpful in many situations.

3.5.2 Desired Attributes

In order to construct a control chart applicable to the start-up situation, the requirements and assumptions must be considered first. For most production environments, the eventual outcome of process data is traditional SPC charts, such as individual response and moving range charts. This is important for two reasons: first, the assumptions used for control charts should be applied to the interim solution and second, the ability to switch from the interim solution to the traditional control charts becomes a complication. The design of an interim solution must address both of these issues.

It is important to recognize that there are several different phases to the typical start-up cycle which demand different requirements from a control scheme. Before a full-scale manufacturing start-up, there is a development effort which might create a product prototype, undertake some degree of process modeling, and define product attributes. Then, there is a transition period between development and manufacturing which includes technology transfer and some development support during limited manufacturing. This begins the very initial phase of what most term "start-up" production. After this, there is typically a ramp-up schedule where goals are set based on the ability of the organization to learn how to produce more on the given equipment until the full production goal is attained. Understanding the process variables that control the product attributes and tracking them with SPC is most effective during the ramp-up period. This way, manufacturing personnel have an introduction to the strengths and weaknesses of the process itself.

Figure 20: Start-up manufacturing phases



The most obvious hurdle to overcome in a startup situation is the dearth of data. Typically, between 20 and 40 subgroups of data are collected before the mean and control limits are calculated[22]. Unfortunately, the data available before production exists from design prototypes, or worse yet, estimates. Both of these are inadequate for establishing control limits or realistic production targets. So, an interim control charting method must recognize the lack of historical perspective. One aside, while design prototypes do not provide a proper historical reference for production control charts, they can provide valuable information about production fixtures. Variability takes many forms in production, one of which is measurement variability. Using design prototypes to establish measurement capability is an important effort before production begins. By running a few prototypes through the measurement equipment multiples times, each fixture's variability can be defined. Additionally, as mentioned in the Background section, inclusion of prototype data increases the number of historical data points available for control limit calculations. As long as the manufacturing process is close to the process used to make the prototype, product attribute data can be used initially for control limit calculations.

Another desired attribute of an interim control charting method would be the ability to detect process changes quickly. Because there are likely many process changes occurring daily, gaining feedback on the effect of these changes is helpful for future rounds of decisions. This need can be reflected in the sample size for an observation. In continuous processes, this would be reflected in the polling time between measurements – in specific, shorter times between measurements allow for more feedback sensitivity. In discrete processes, an analogous result comes from using smaller subsets to plot data. One caution is that by increasing sensitivity, many times the false alarm rate also increases. This trade-off needs to be assessed especially if the control chart method is similar to the long term display. If the false alarm rate is too high, the manufacturing personnel may learn to ignore the signals in the future.

One of the most useful features of control charts are the control limits themselves. They provide a guide for what the process is capable of as well as a reference point for choosing targets and external specifications. Control limits also provide a historical context of the process itself – in essence, showing the range in which "good" product is produced. Unfortunately, control limits require the most data for accuracy and are prone to error otherwise. Additionally, without thorough process knowledge, control limits established by the data mean and variance have little practical meaning when establishing a target production range. A start-up interim solution should

[22] Rocke.

include the benefits of control limits such as giving historical context, while overcoming the difficulties they involve.

One of the final preferences for an interim charting solution is for it to resemble the long term statistical process control solution. Using this as a secondary goal allows minimal additional operator and technician training, as well as simplifying the IT for the SPC displays. This will become more important as product lifecycles shorten and production personnel are subjected to additional change as products increase in complexity.

3.5.3 An Interim Solution

The first challenge to start-up control charting is the lack of data, as mentioned previously. While there is not a solution besides time, this fact dictates the first attribute of an interim solution. While traditional control charting prefers larger subsets of data, in a start-up situation this is not possible. Therefore, subsets of data should be smaller, even using each individual response, initially. There is basis for this decision past the necessity of the start-up situation. Because there is no history of process capability, each new observation is significant by itself. Secondly, because production volume is generally lower during the start-up phase, the data points are more independent from the previous and subsequent measurements as changes to the environment are more likely in between units produced. Further, these changes are made so frequently in the beginning of the process, the chart must be able to detect the effects immediately and not damper the response by averaging it with other observations. Based on these reasons, an interim control chart subgroups must have small subsets initially. If there is doubt about how many data points are independent, a good initial value to use is one until enough evidence is collected to convincingly suggest another number.

Luckily, this is nothing new for control charts – individual observation and range charts are common in industry. However, these traditional individual observation charts still have the disadvantage of requiring upper and lower control limits based on historical data, something that a production start-up does not have. Instead, the small data set itself must give an indication of what the process is capable of at the given time. So, if the area is striving for six sigma performance, the control limits can be generated based on data set itself. This is where more detailed information about the specific process is required. In most cases, by the time SPC is requested, at least the first twenty units have been produced. If an individual response chart is appropriate during the initial production ramp-up period, this is enough data to calculate control limits with some statistical significance.

Using a small number of data points to find the control limits is out of necessity in most cases. However, if we take the ramp-up periods to be sets of smaller start-up periods, we can extend the notion of using less data to determine how the process is progressing. As stated earlier, because process changes are so abundant during the start-up of a new manufacturing process, control limits are not necessarily valid if calculated cumulatively over all production data. This is especially the case if control limits are found for process conditions. In the case where we are treating each data points as approximately independent, then it becomes feasible that possibly only the previous twenty or thirty data points might have similar enough process conditions to be related. Using this idea, an interim solution can track changes in the process as well as give guidance on the performance by dynamically calculating the control limits based on the most recent data.

3.5.4 Examples

In this example, the process is viewed with the typical Xbar chart with subgroup equal to five, an individual response chart and using the interim method of charting. The first distinction to notice is between the typical Xbar chart and its individual response partner. Immediately, it is obvious that the process response is dampened by averaging over subgroups, especially in the first part of the data set. Almost immediately the Xbar option can be discounted for start-up production use simply based on its inability to detect small shifts quickly.

Figure 21: Variable Control Chart, XBar using subsets of five[23]

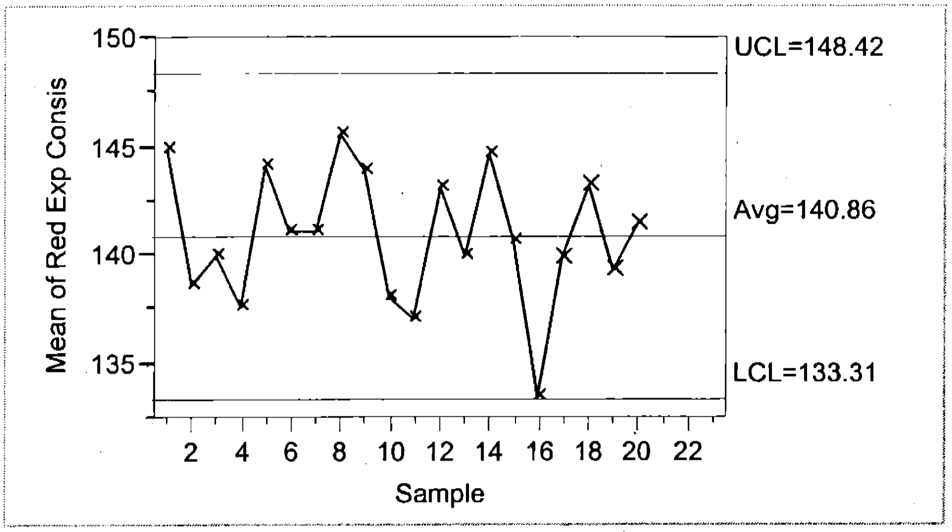
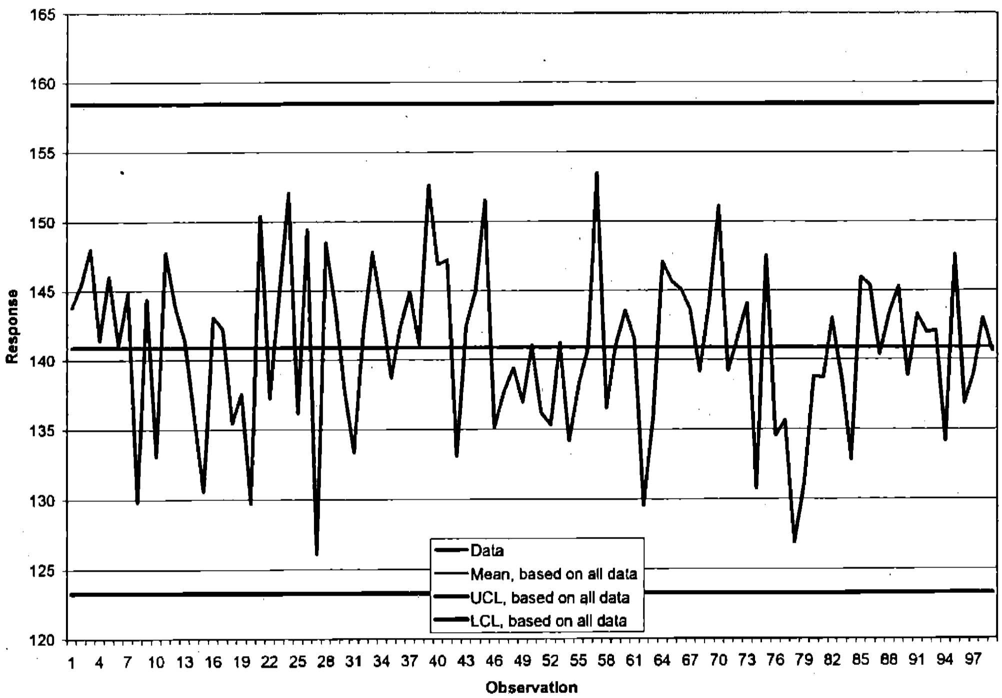
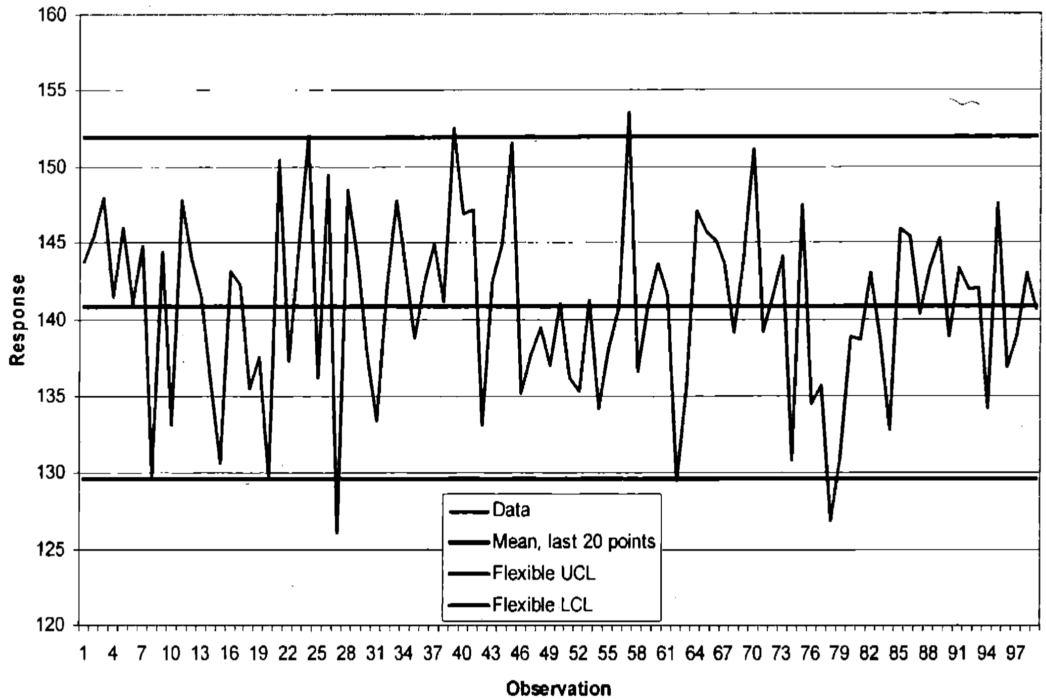


Figure 22: Control Chart, Individual Measurement of Variable 1



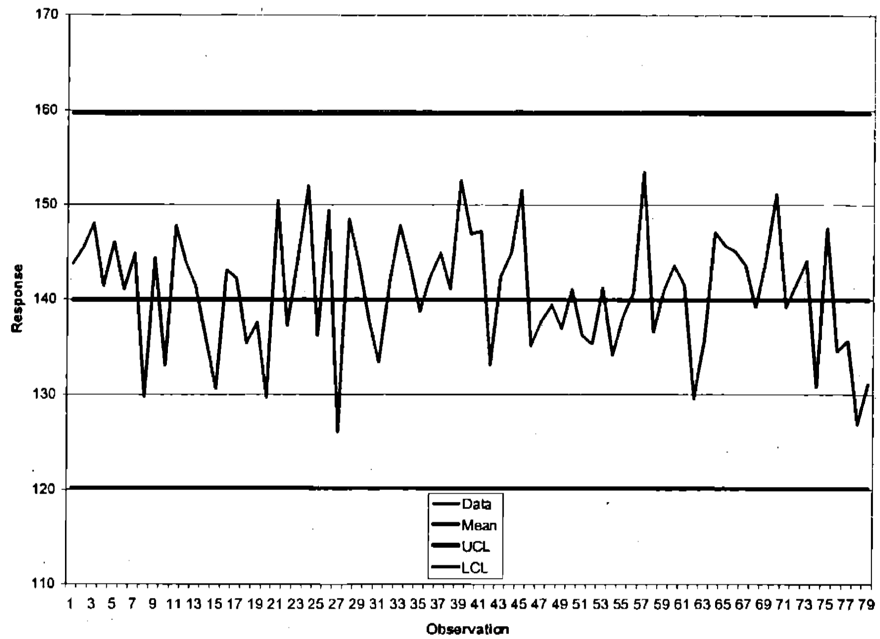
[23] Output from JMP Software, SAS Institute.

Figure 23: Interim Solution Control Chart



Once the choice is made between the subgroups, the next step is to place control limits onto the graph in order to see what guidance it provides. The above Interim Solution chart uses "flexible", or dynamically calculated, control limits to display where the most recent data has shown capability. The above chart shows that the most recent data points show an improvement in variability over historical data, while maintaining a similar target value. Additionally, while basing the control limits on all of the hundred data points shows no points which are outside of three sigma of variability, the control limits based on the most recent data would consider these previous abnormal. Plotting the first eighty points only shows that the most recent data points in that case (61 – 80) are consistent with the first 60 data points. Comparing this with Figure 23 shows clearly that points 81 through 100 demonstrate a significant process change which should be investigated.

Figure 24: Interim control chart with dynamically calculated control limit on the first 80 data points



An additional area of benefit from the Interim Solution chart comes when there are targets or specifications set before production begins. This scenario is common when the new product is similar to a previous version or when a customer has critical specifications that must be acheived. Again, traditional control charting treat control limits as stationary entities at all points in time. For cases of production with existing targets/specifications, many companies set their control limits before any data has come available as targets for the future. By using the dynamically calculated control limits based on recent data, these targets gain more significance as they are visually compared with most recent production data.

Figure 25: Interim Control Chart with Specifications, all 100 data points

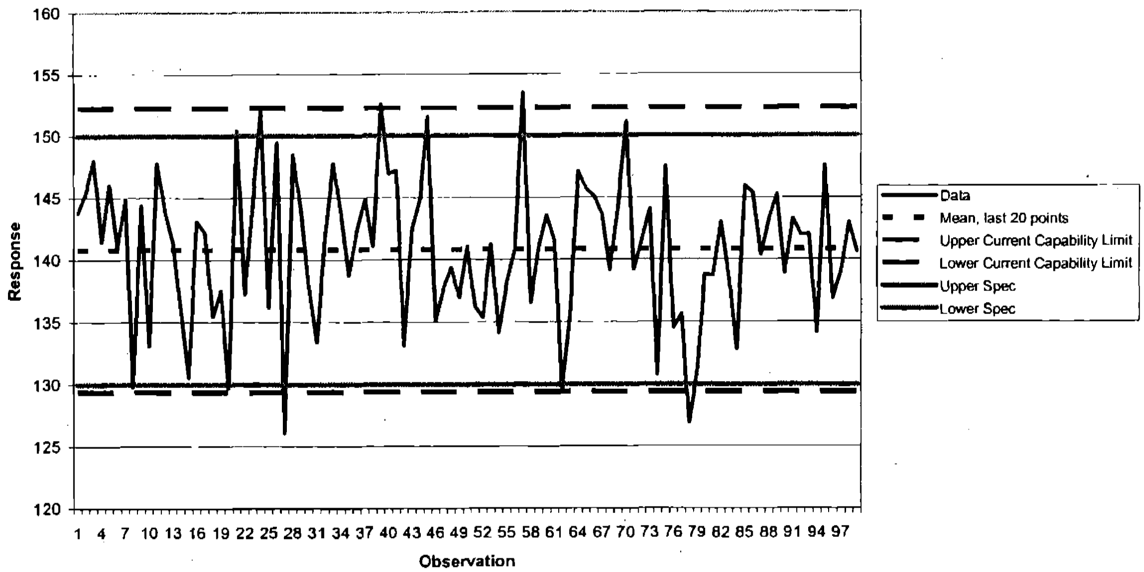
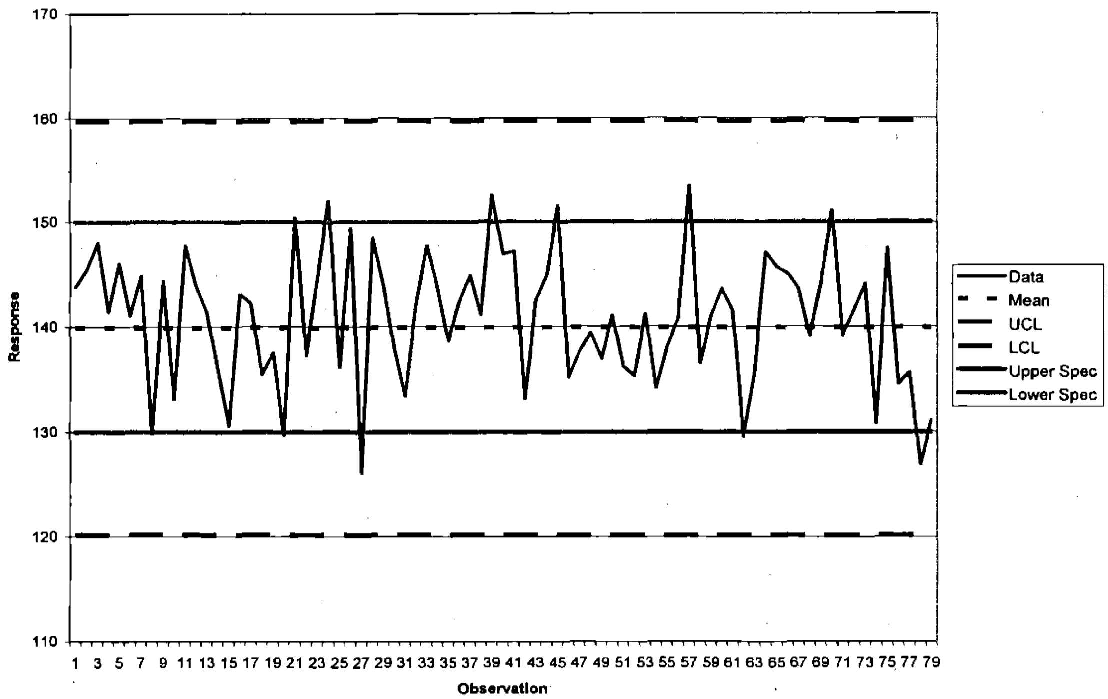


Figure 26: Interim Control Chart with Specifications, first 80 data points



From the previous two figures, it is observed that with the first eighty units, the production data is in control (with respect to the red lines) but its capability to meet the specifications (green lines) is questionable. This is in contrast to chart including the next twenty data points where the control limits based on the recent twenty data points show that the production targets are close to being met but the first eighty data points were not consistent with recent process improvements. Because the calculation of control limits based on only the most recent data points is a guide to the capability the process is showing under current conditions, these control limits are better termed Current Capability Limits, or CCLs.

The particular strength of using the most recent data points only for the calculation of the control limits is that overall variability is being tested for any process changes occurring. Other charts, such as CUSUM and EMWA are more adept at finding directional shifts in the data, but overall variability is challenging to observe from these charts. There are some drawbacks, however, to this technique. The assumption that smaller subsets of data are independent needs to be validated before any action is taken based on a chart alarm. If the data it not independent from each other many false alarms from typical process changes such as a new lot of raw material or shift change will result. Secondly, basing control limits exclusively on the limited data makes them less accurate. Depending on the sensitivity of the final product specifications, this could be a significant negative result.

3.5.5 Including Robust Techniques within Interim Solution

A slight modification to the above technique is useful in situations where there are extreme outliers present regularly in the data. Many processes produce outliers or data values that are inconsistent with normal process conditions, especially if the measurement fixture has error values it sends through the IT system in response to a known run condition. In these cases, using the median of the recent subgroups or data points instead of the mean makes the control chart more robust to these outliers.

An additional function of the robust control charting is to remove obvious outlier points within the data sets. One method applicable to the start scenario is the use of X_Q and R_Q charts. These chart types estimate the process standard deviation by using the interquartile ranges (IQR) of the data instead. Because the control limits are less affected by outliers, data points indicating special causes are detected more efficiently than if the outlier data points were included within the calculation of the control limits.

3.5.6 Summary

In summary, charting initial process data in order to visualize differences in overall variability requires the following:

1. At least twenty data points for statistical significance.
2. Small subsets, even individual responses, make process shifts visible faster.
3. If basing control limits on y data points, plot at least 2*y data points to gain historical perspective.
4. Inclusion of robust charting techniques increases the detection of special cause situations.
5. Use of targets or specifications is optional but greatly benefits knowledge of future improvement efforts required to achieve those targets.

Once the ramp-up to new production is complete, transition to traditional SPC charting is necessary. The above technique provides a useful display of data during the start-up period of production by does not offer the accuracy necessary for long term improvements. Instead, the process for transitioning to traditional SPC charts should include the following steps:

1. Collect the proper amount of data to capture all known sources of variability such as different shifts of manufacturing personnel, lots of raw materials, and performance ranges of manufacturing equipment.
2. Perform a Gage R & R to understand the measurement capability of the manufacturing equipment providing data about the process.
3. Calculate the UCL and LCL based on as much process valid process data as possible.
4. Establish the appropriate run rules to capture inconsistent lots of product while not creating a significant number of false alarms.
5. Understand when process changes are significant enough to warrant another iteration of these efforts.

Using these techniques once the process has stabilized will ensure that consistent product leaves the process long after the start-up phase is completed.

This page intentionally left blank.

4.0 Case Example: Kodak Digital Camera

4.1 Background

Any new product iteration is difficult to model, but the Kodak Professional DCS Pro 14N was especially challenging. The product itself was an amazing leap in technology, doubling resolution for roughly half the price of any competitor's product on the market. This was to be a product that blurred the line between professional photographers and advanced amateurs. Even though it was positioned to be a commercial success, without adequate process modeling and control the DCS Pro 14N manufacturing team would have problems delivering the camera as specified by the development team.

In addition to new technology within the camera, completely different processing equipment for several critical steps in the manufacturing cycle was introduced. To further complicate the task, many new vendors were being used to source components for the camera. While process data for the components could include up to ten additional companies, it was decided that the Kodak process itself would be the first analyzed. Kodak's portion of the process could best be described as a final assembly, calibration and test area for the DCS Pro 14N. Because this was a discrete process, data collected from each of the process stages was mapped with unique identifiers in order to compare data accurately for each unit.

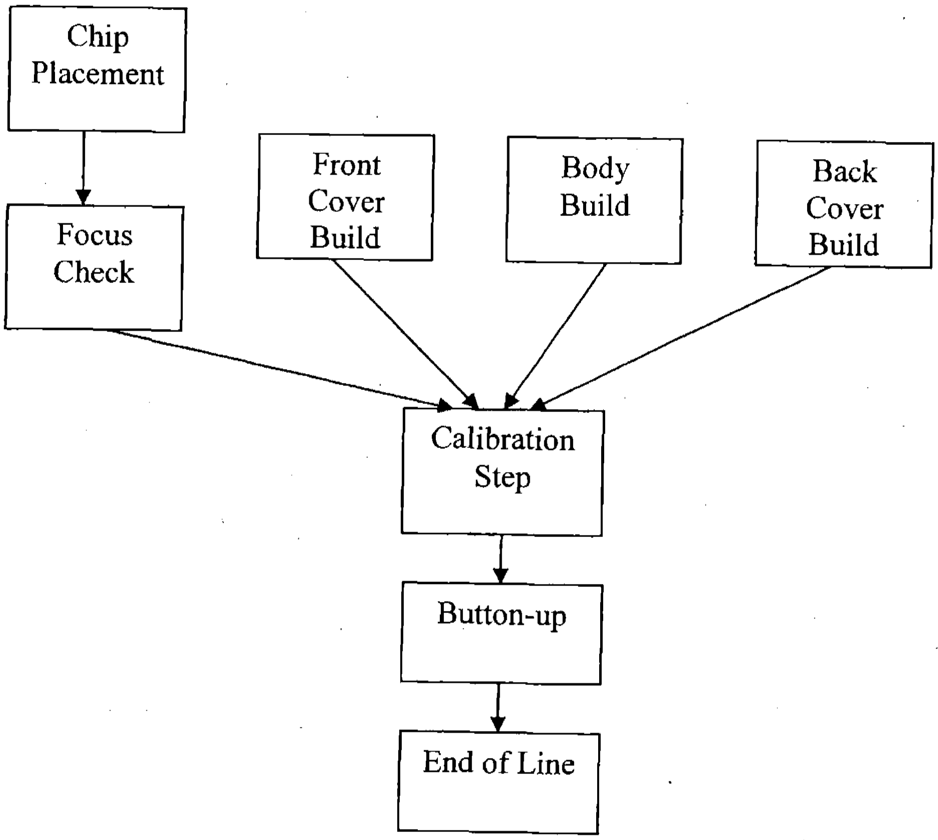
The DCS Pro 14N was significantly different, removing the possibility of building a process model from historical data. The only exception to this was the final quality inspection step, where many of the customer limits were well defined. So, not unlike many other process start-ups, this product had very clearly defined final specifications but not generally quantified manufacturing specifications to meet the final specifications. However, during the analysis it became apparent that many of the previous final specification were artificially tight due to higher control in the previous camera programs. As the new process came online, final specifications were altered based on substantial evidence of acceptability to ensure high-quality product flow to the customer.

Digital camera design is centered on the imager chip. An imager chip is an electrical device which, upon shutter actuation, measures light intensities through a filter. The information goes through integrated software to optimize the picture based on the particular imager chip installed. Then, a digital image is rendered on an LCD screen on the back cover of the camera. The camera body itself is comprised of three large components. The camera front includes the fixture for different lens mountings and some minor electronic components. The imager chip is housed inside the camera body, along with the shutter and auto-focus mechanism. The camera body also includes the majority of the electronic components. Finally, the camera back acts as an interface with the user with its LCD screen and buttons for operation. Positioned between the camera front and back, the camera body acts as a controller for the peripherals housed in the other two components.

Customer requirements are easy to understand but difficult to quantify. Images rendered from the imager chip need to appear as true to reality as possible. While the human eye can quickly distinguish, programming software to understand the subtleties of sight is a large undertaking. To simplify the explanation, we can break the picture quality down into two major components, focus and image quality. Focus is the crispness of the image. Image quality includes many subcategories, but can be described simply as the color integrity of the picture.

The process itself was relatively straightforward. Raw materials entered the process at many different stages, but process parameters were measured at only four. The first major stage positioned the imager chip onto a stiff plate based on focus optimization algorithms. Then, an independent measurement verified that the placement was within tolerances set by customer requirements. In parallel, the subcomponents were built with the binary requirement of sending a charge. All the components met again at the calibration stage. This was the stage containing the most Kodak technology as any flaws on the imager chips were "mapped" out using software installed on each camera. This calibration step was the bottleneck of the process due to the plethora of image quality tests used to insure a realistic image. Finally, after some minor cosmetic additions to the camera (named "Button-up"), each unit was analyzed for quality at the "End of Line" station. While the calibration steps worked to map out imager chip defects, the End of Line tested the success of this step by using the camera in a customer space approximation.

Figure 27: Process Outline



The calibration step generated thousands of process attributes for each unit. This data was captured at various stages of completion, so as the chip continued in the calibration process, the data changed. Additionally, the bottleneck step was limited by high fixed cost fixtures that were custom built specifically for the process. This meant that process analysis involving the bottleneck step would be highly complicated and involved, yet held high potential for efficiency gains.

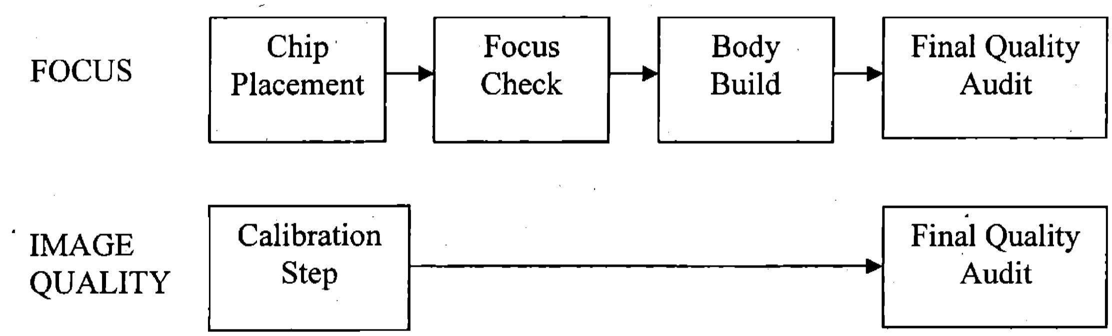
4.2 Data Generated

The large amount of data per camera unit alluded to earlier would present a challenging task for any analyst. Fortunately, it was assumed correctly that many of the process attributes were either redundant or had little to no effect on the final product. So instead of becoming an extremely complicated modeling case, Kodak's modeling effort focused more on which data should be tracked in order to maintain control of the product quality.

To begin the modeling process, the first step was to acknowledge the sources of data. Since this was a new production line, the data collection systems were updated within six months of the start of production. This allowed for proactive data structures to be in place as the product start-up occurred, facilitating "on-demand" data analysis. Major data generating steps were the Chip Placement, Focus Check, Calibration Step, and End of Line stages. The other stages simply functioned as build areas where the product either was correct or not.

Next, understanding the types of data generated at each fixture helps to simplify the analysis process. The camera process attributes were separated into two major categories: focus and image quality. For this process, the majority of the focus attributes were captured during the Chip Placement and Focus Check while the image quality attributes came from the Calibration and End of Line steps.

Figure 28: Diagram of data types and their related process stages



Interestingly enough, even though the larger amount of data was generated for the image quality category, the focus category went through more physical process steps. Analyzing the focus attributes provided a clear example of benign, new attribute, unintentionally altered and intentionally altered process steps as well as the discovery of necessary data that was not collected at production start-up. The Chip Placement stage represents an intentionally altered process step for Focus. This is the stage where the chip receives treatment to improve the desired attribute, focus. Additionally, this step also acts as an unintentionally altered process step for the Image Quality attributes. As the chip is exposed to the elements, it can be scratched or pick up particulates from the air; both affect image quality. The Focus Check would be considered a benign step for Focus as the measurement itself does not affect any attributes of focus. The body build step, where the imager chip is placed inside the camera body, is an unintentionally altered step for focus. This stage is designed to assemble a camera body, but the imager chips can be jarred, thereby affecting focus. This was not known initially and there was no data systematically collected to understand the shifts caused at this step. It took troubleshooting analysis of the available data to narrow down the problem to that particular step. Finally, the End of Line step is simply a measurement of the image quality and focus attributes. Because it is a measurement and involves little to no interaction with the underlying electronics, it can be considered a benign process step.

This camera build process provides an example of how to approach a process mapping effort before data analysis. First, a simple chart displaying the flow of camera parts (Figure 27) helps to understand the possible areas where the process can affect the product attributes. Then, by separating the important (as defined by customer requirements) product attributes and independently analysis the process steps designed to alter the product attributes an analyst gains a lead on possible correlations important to the final outcome. As an additional outcome, these process maps can be referred to when alterations are found in the product which cannot be accounted for within the steps designed to modify the product. For these reasons, a detailed process map is a critical first step in and process analysis or control effort.

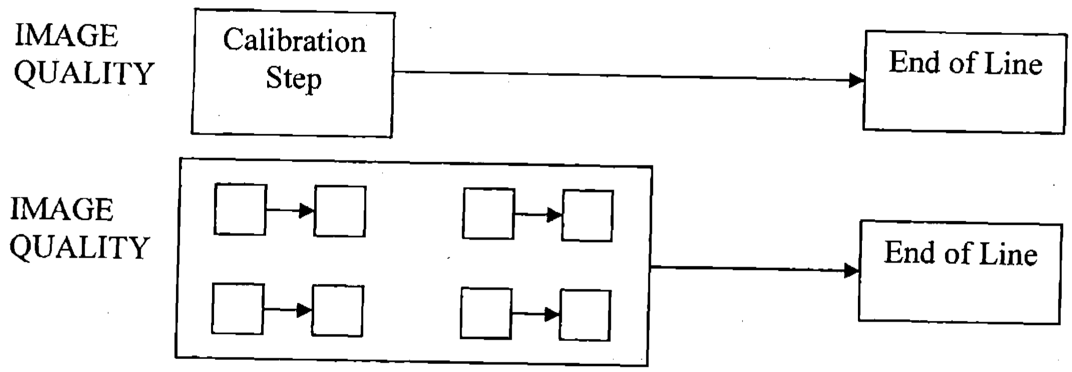
4.3 Data Analysis

The first step in data analysis is to decide on a thorough plan of attack. In Kodak's case, the high demand for the product and mounting backorders forced the bottleneck step to be considered first. In this case, as mentioned earlier, the bottleneck step is the calibration where thousands of image quality process attributes are generated. Overall, this appears to be a simple case, as there are two associated process steps, Calibration and End of Line. However, the complexity of the data and shear volume of attributes warrants a more detailed look.

The calibration step is subdivided into several smaller steps. By its very nature a "calibration" step accepts some starting level information and then is designed to make alterations to force the product to some "ideal" value or range of values. Therefore, some of the steps are designed to measure data before changes are made to the electronics and software and some attributes are generated to assess whether the changes were successful. This leads to a collection of data points that can be labeled as either pre-process or post-process.

Because of the complexity inherent at this step, it is important that any analysis attempted is done with the understanding of the underlying process. In a production start-up case when time is limited, definition of all sub-process performance levels was too detailed. Instead, a first pass model using the "final" state numbers – the data measured after each calibration alteration – identified which areas held the most promise for improvement. Additionally, because conditions during manufacturing contribute to product variability, it is important to include them into modeling early to ensure that correlations have true physical meaning. Also, including condition data early in the process can identify which condition data is lacking, as process attributes that do not correlate with process conditions are rare.

Figure 29: There are a range of sub-processes within the Calibration process step itself.



4.4 Determination of Critical Process Parameters Example

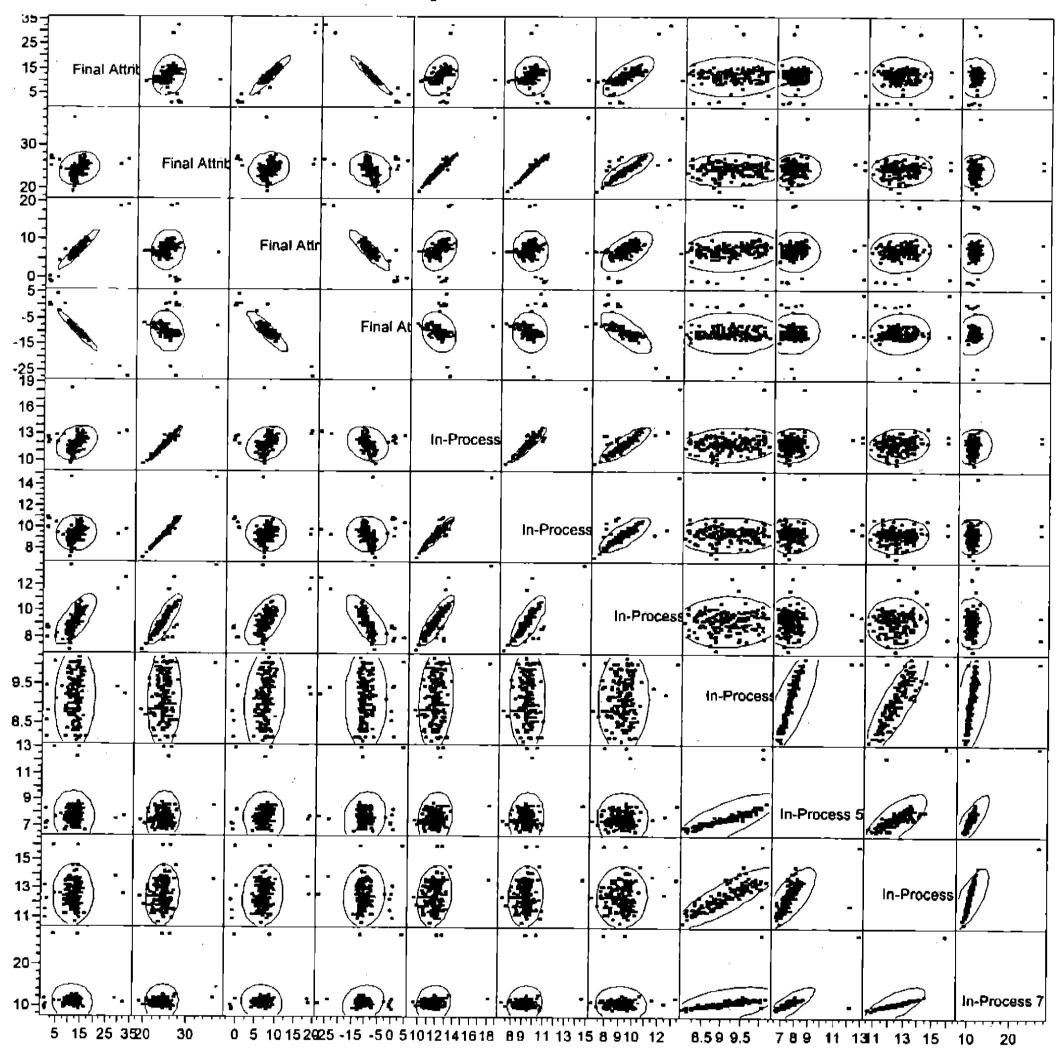
As an illustration of how to use process modeling and interim process control charting during a start-up situation, product attributes and process conditions from the image quality portion of Kodak's data will be analyzed. Because of the strong relationship between image quality and the customer's perception of the overall product's quality as well as the bottleneck nature of the calibration step, this portion of the process became highest priority for analysis. The issues revolved around setting in-process specifications that ensured that final specifications were met. But, before this could occur, the relationships between the process conditions and final product attributes needed to be established before bounds could be assessed.

In this example, there are seven in-process conditions and four final product attributes that are all related to the same aspect of image quality. The assessment begins with a single-stage analysis which determines which variables from the calibration and final quality inspection are independent and therefore ascend to the next level of analysis. Then, these PSCVs are compared with each other to understand the possible relationships or lack of explanatory variables for the final product attributes.

Step 1: Initial Scatterplot Screening

The first activity in this type of analysis is usually a simple scatterplot of all the variables to observe if there are any correlations at all. In this case, there appears to be some redundancies within the single steps as well as some correlations between the in-process and final attribute variables.

Figure 30: Scatterplot matrix (bivariate relationships) of the four final product attributes and seven in-process variables[24]



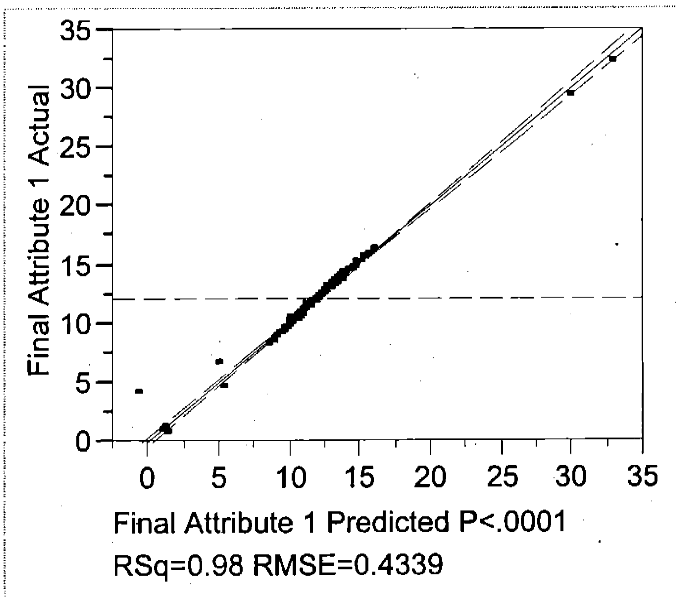
[24] Generated by JMP Software, SAS Institute.

With the scatterplot showing promise of correlations, the most detailed analysis can begin.

Step 2: Single Step Analysis

The point of a single step analysis, as outlined earlier, is to reduce redundancy in the process variables requiring observation and further analysis. For this example, there are two effective stages, a calibration step with variables labeled "In-Process" 1 through 7 and a final quality inspection step with variables labeled "Final Attribute" 1 through 4. Beginning with the last step, the final attributes are compared with each other to better understand the underlying correlations. Using basic statistical software, the first variable is modeled by the remaining three final attribute variables. If the fit is strong enough, the entries are made into the DSM process variable matrix. However, if the minimum fit criterion (this analysis used an $R^2 \geq 0.8$) is not met, then the variable is considered independent of the other variables within the process step and moves onto the next level of analysis.

Figure 31: Fit of Final Attribute 1 with Final Attribute 4[25]



With each iteration, two activities are required to ensure that important data is not discarded. First, any outliers should be noted and compared with any known critical product failures. If there is a relationship between the outliers of a given correlation and some product failure, then that process relationship itself is critical to monitor. Secondly, the residuals from each model require

[25] Output from JMP Software, SAS Institute.

analysis to ensure that the model follows the assumption of random variation and normally distributed residuals. If this requirement is not fulfilled, the model will likely be unreliable.

Step 3: Variable Reduction using Correlation Matrices DSM

Taking the information from the Single Stage Analysis and using the DSM methodology, the following matrices result:

Figure 32: Summary Process Correlation Matrices

	Observations				
Model	Final Attribute 1	Final Attribute 2	Final Attribute 3	Final Attribute 4	Row Sums
Final Attribute 1				1	1
Final Attribute 2					0
Final Attribute 3	1				1
Final Attribute 4	1				1
Column Sums	2	0	0	1	

	Observations							
Model	In-Process 1	In-Process 2	In-Process 3	In-Process 4	In-Process 5	In-Process 6	In-Process 7	
In-Process 1		1						1
In-Process 2	1		1					2
In-Process 3		1						1
In-Process 4					1	1		2
In-Process 5				1		1	1	3
In-Process 6							1	1
In-Process 7				1	1	1		3
	1	2	1	2	2	3	2	

The gray highlighted cells on the in-process matrix represent an interaction that is significant to the predictive model for in-process variable 4. After reducing the matrices using the procedure given previously, it becomes apparent that final attributes 3 and 4 along with in-process variables 1, 3, and 6 can be expressed by the other variables, thereby making them redundant within the next level of analysis.

Figure 33: Reduced Process Correlation Matrices

	Observations		
Model	Final Attribute 3	Final Attribute 4	Row Sums
Final Attribute 3			0
Final Attribute 4			0
Column Sums	0	0	

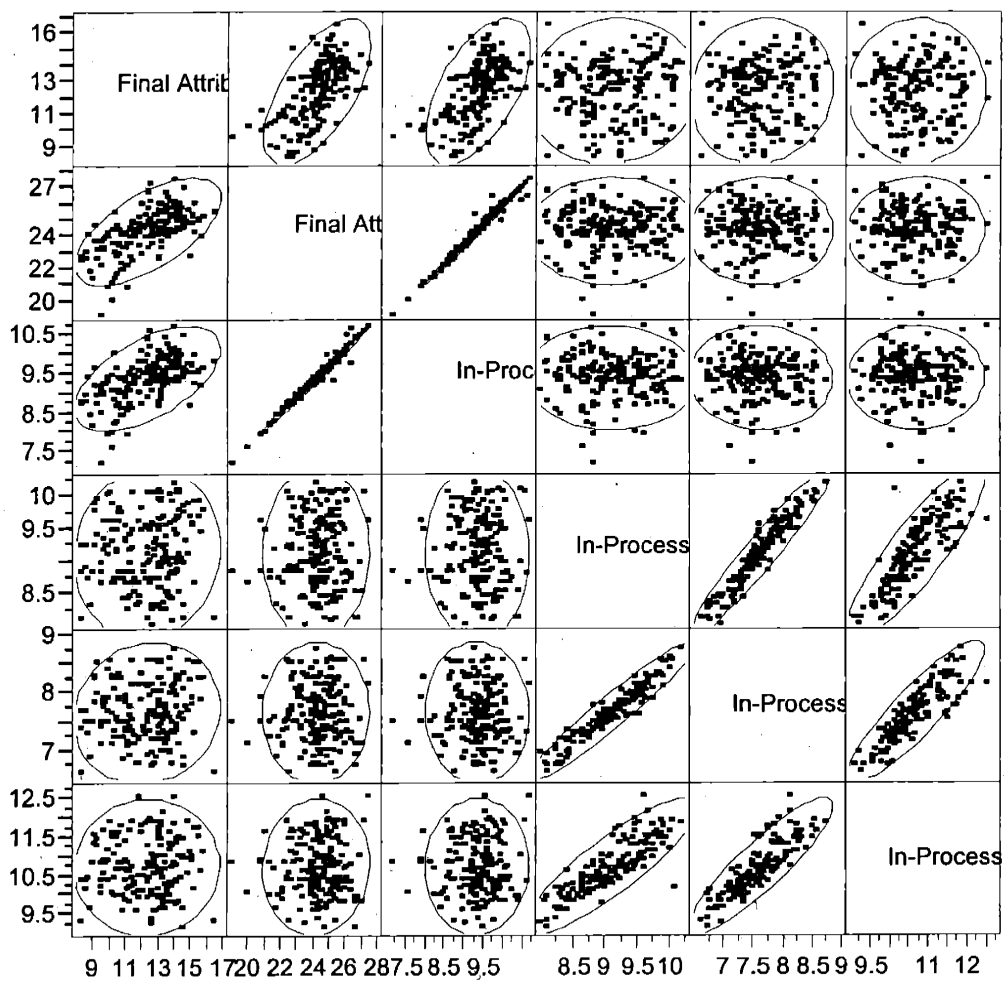
	Observations			
Model	In-Process 1	In-Process 3	In-Process 6	
In-Process 1				0
In-Process 3				0
In-Process 6				0
	0	0	0	

Additionally, using the PLS principles, because these variables can be expressed with linear relationships of the other variables, as long as final attributes 1/2 and in-process variables 1/3/6 are monitored, the redundant process variables are tracked inherently.

Step 4: Sequential Step Analysis

Using the variables from the single stage analysis, the next level of modeling considers correlations between the in-process variables and the final attribute data. Sequential step analysis is identical to Step 1 and 2, leading the analysis to begin with a scatterplot to quickly observe if correlations are likely. In this case, one can observe a strong correlation between one of the in-process variables and final attributes.

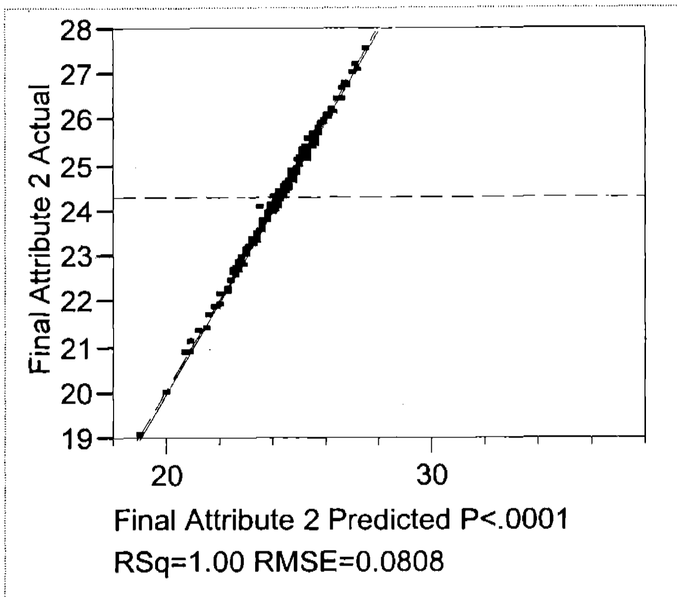
Figure 34: Scatterplot of bivariate relationships present in the sequential stage analysis[26]



[26] Output from JMP Software, SAS Institute.

Next, using a typical statistical analysis tool, each variable can be assessed for promising predictive models. Again, using the minimum acceptable R^2 value of 0.8, a relationship is discovered that predicts final attribute 2 using final attribute 1 and in-process 2.

Figure 35: A strong correlation exists for final attribute 2 and a linear combination of final attribute 1 and in-process 2



Because, no other correlations meet the minimum criteria, the DSM methodology does not need to be applied. Therefore, after the sequential stage analysis, we have the following variables to track:

- Final Attribute 1
- In-Process 2
- In-Process 4
- In-Process 5
- In-Process 7

Overall, this analysis reduced the number of system variables from twelve to five. Additionally, it has shown that the in-process variables provided do not create a complete set of explanatory variables for the final attribute data. Ideally, all final product attributes should be a function of upstream variables. This way, product quality can be controlled upstream where there is lower value to the work in progress. Further research into in-process variables is warranted in this case.

Step 5: Use Critical Variables for Interim SPC Charting

Until the full process modeling has been completed, the knowledge of the critical variables can be used to track process variability in the interim. For instance, looking at the most recent 100 data points of Final Attribute 1 using both the traditional control limits and the interim capability limits displays different pictures.

Figure 36: Last 100 Final Attribute Data Points using Traditional Control Limits

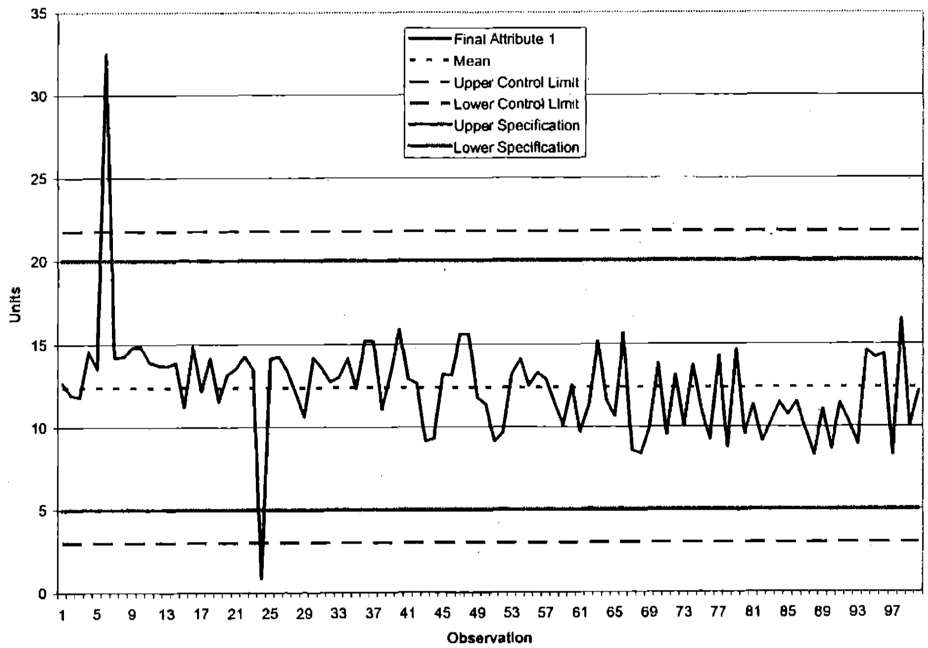
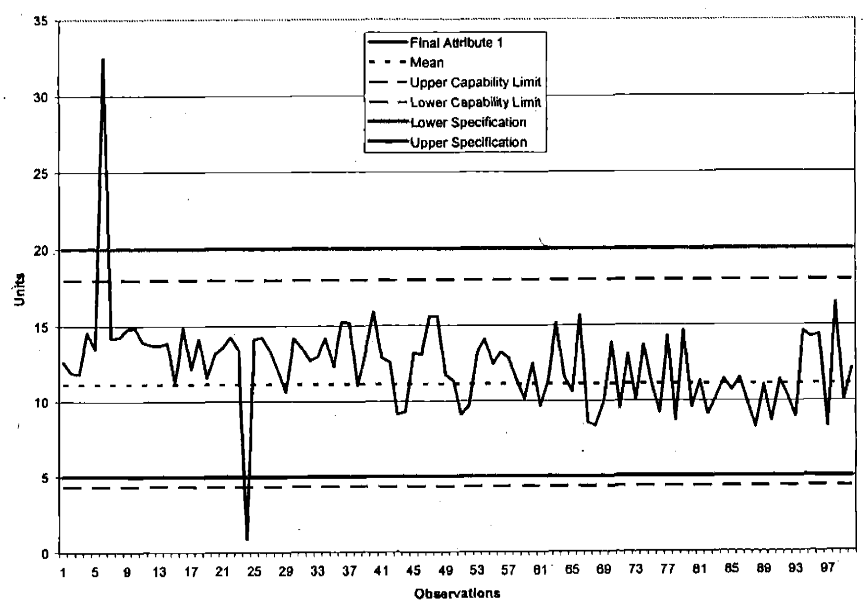


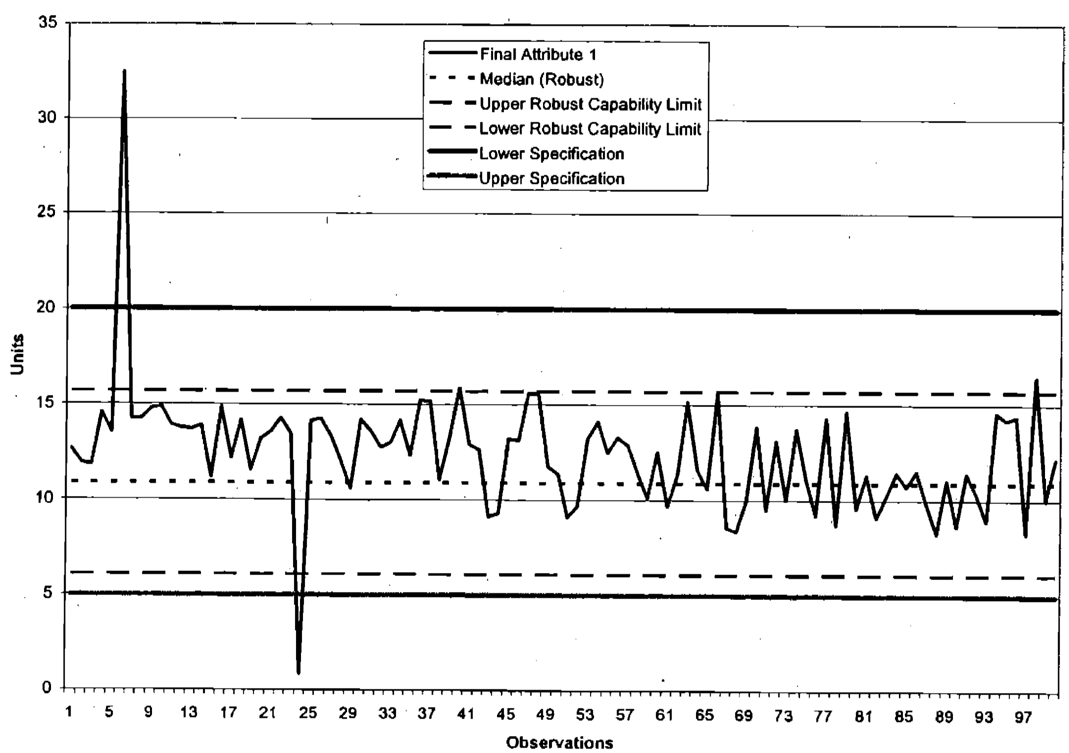
Figure 37: Last 100 Final Attribute Data Points using Capability Limits



Using the traditional SPC plot, we see that the control limits are outside of the desired specification limits. Additionally, even though the data is slowing trending downward, this is not readily apparent by viewing Figure 36. However, using the interim control chart during these first 100 units, we see that the last twenty data points establish that the process is nearly capable and that the mean of the most recent data is different from the first portion of data.

Applying robust control charting techniques (Figure 38) in addition to the interim chart philosophy yields a slightly different outcome. Using the most recent twenty data points, similar to Figure 37, the median and interquartile approximations of standard deviation exhibit slightly tighter process capability limits. This chart allows roughly equivalent observations such as the recent data having an expected value lower than the earlier data points and the process being slightly off-center when compared to the desired specifications. However, the robust control limits are more strict, providing additional information about particular data points needing attention.

Figure 38: Last 100 Final Attribute Data Points using Robust Capability Limits



In both the robust and interim only charts, the capability limits are near to the desired specifications. This shows that at the current process conditions, the attribute being measured will likely fall within specification. During a production start-up this can be valuable information as processing conditions tend to change rapidly.

Using the above techniques allows for efficient ways to plot process data while the more formal process modeling takes place. During a start-up situation where there is not enough data to establish process control limits accurately, these methods provide an effective way to monitor the important process variables.

4.5 Summary

The Kodak case example provides a valuable illustration of a methodology to initiate data analysis within a start-up environment. After mapping the assembly process and the data generated at each step, the critical attributes for the customer can be prioritized above those attributes less noticeable. Then, the first available data from production start-up can be analyzed in a standard procedure starting at each individual step in order to establish those process variables affecting the final product attributes. Finally, by using slight variations on the traditional control charting methods, the critical process variables can be displayed in an effective manner for proactive response.

The short product life cycles of digital cameras force production ramp up periods to reach full capacity in a much abbreviated time frame in order for the individual camera product to be profitable. Application of a thorough process modeling and a control technique is necessary to increase process improvement and quality in the short term. For the long term, this information can be utilized in future camera products to address known design issues or performance weaknesses.

5.0 Conclusions and Further Areas of Study

After applying an interim SPC display for production start-up, there should be noticeable differences in the way daily activities are completed. While some results are easily quantifiable, others are harder to measure, but contain just as much value to the productivity of an organization. Especially during a manufacturing start-up where resources are typically stretched, the less tangible benefits from instituting a data driven culture become critical to a successful outcome. Further, because of shorter product cycles and stronger competition, the ability to affect product quality faster will become increasingly important.

New projects, such as creating the IT structure necessary to implement the above strategies, normally require evidence of positive financial return to the company. Installing SPC and limited process modeling during the startup phase has some immediate benefits. First, instituting SPC charts, even just at the final process step, can minimize the number of out of specification product released to the customer. Secondly, the problem solving process is shortened tremendously with the access to process data and trends. Finally, by determining the redundant process parameters early, the necessary hardware can be decreased.

The larger rewards from installing SPC tend to result from the increased frequency and quality of interactions between team members on the production floor, engineers, and designers. Because the data displayed can be viewed by many more individuals, there is an increased likelihood of catching problems earlier. Also, because the different departments see the same data, the communication contains common information. This type of transparency within an organization must be transitioned slowly if new to the area. In many company cultures, individuals feel as though databases are a way for others to interfere with their job role. To combat this issue, the rollout of a new data display and access system must include all areas together to show impartiality.

Another more subtle advantage is the improvements available to the next product cycle. Product launches are an extremely delicate period that strongly determines success for the company's efforts. If demand cannot be met and manufacturing ramp-up is slow, there are opportunities for competitors to enter or customers to find substitutes. Because information during the start-up period is sometimes lost once the processes are fully ramped up and producing adequate product it is seldom reviewed for the next design cycle. By knowing how to reach full manufacturing potential more efficiently, many launch problems can be mitigated and improved for the next product cycle.

In totality, efforts to understand the process through variable correlations and interim SPC shortens ramp-up and design time of the next product, in addition to the tangible quality benefits common with traditional SPC implementation.

Further Areas of Study

Because the combination of SPC, DSM, and PLS techniques use the most basic level of each method, there is opportunity for improvement. Within the SPC portion of the tools, there are hundreds of options for control schemes and different charting displays that benefit responsiveness and accuracy. Additionally, with the power and programming ease available in computer algorithms, it is possible to create criteria that signal when the process data is consistent enough to warrant more accurate control.

The DSM methodology used in this thesis is of the most basic type: binary entries. Recent efforts in DSM have moved closer in the spectrum to PLS. By making the entries in the DSM reflect true correlation strengths, it approaches the PLS technique. Therefore, if interactions between process variables can be compared on the same basis from the beginning of start-up, a more detailed matrix can be created from the initial analysis. This would ensure that the best models are used and redundant variables truly have little impact on the final product attributes. Additionally, PLS analysis can be programmed to be completed automatically given new sets of data. It is feasible that difference in the PLS/DSM matrix can be compared between each modeling iteration to reveal when the process has reached an equilibrium point.

Another technique not explored in this work, but containing significant possibility is Design of Experiments. The modeling techniques described above are based on the assumption that process variables cover a broad range during the initial phase of manufacturing. This may be a flawed assumption, however. A true design of experiments tests the full range of the processing equipment to determine correlations with more accuracy thereby allowing better prediction of response and possibly more efficient run conditions. If a design of experiments can be carried out before start-up or within the start-up period, the process model will benefit from this higher accuracy.

Management of the design and manufacturing interactions also could further the efforts to lessen start-up production difficulties. By using DSM in its traditional role of outlining interactions between teams, design and early manufacturing cycle can be shortened. Overlaying this with the work required to establish data systems necessary for SPC, both interim and traditional, can better communicate when common data can be communicated across organizations.

Finally, in certain circumstances, financial information can be included in process modeling so that optimization can be based on factors such as maintenance minimization as well as product cost. This type of analysis can have large impact in production environments where fixed costs are large in comparison to marginal costs. As an example, catalyst beds used in chemical manufacturing have lifecycles correlated with temperature ranges and times at those temperatures. If this information is included early in the control design, costs for the production can decrease dramatically.

Overall, manufacturing start-ups are complex events which stretch resources capacity. If tools are planned to model process interactions early, better control schemes can aid ramp-up efforts. By using variants on commonly used production tools such as SPC, DSM, and PLS, transition effort from start-up to full scale production can be minimized. These benefits far outweigh the difficulties they prevent and can be improved and modified based on the specific process in question.

This page intentionally left blank.

Bibliography

Alwan, Layth C. Statistical Process Analysis. Madison, WI: McGraw-Hill. 2000. ISBN: 0-256-11939-2.

Besterfield, Dale H. Quality Control. Upper Saddle River, NJ: Prentice Hall, 2001.

Brassard, Michael ed. The Memory Jogger: A Pocket Guide of Tools for Continuous Improvement. Boston, MA: GOAL/QPC. 1991. ISBN: 1-879364-03-4.

Czitrom, Veronica and Patrick D. Spagon. Statistical Case Studies for Industrial Process Improvement. Alexandria, VA: ASA. 1997. ISBN: 0-89871-394-3.

Griffith, Gary K. Statistical Process Control Methods for Long and Short Runs. Milwaukee, WI: ASQC Quality Press. 1989. ISBN: 0-87389-063-9.

Hawkins, Douglas M. and David H. Olwell. Cumulative Sum Charts and Charting for Quality Improvement. New York: Springer, 1998.

Hayes, Robert H., Wheelwright, Steven C. and Kim B. Clark. Dynamic Manufacturing. New York: The Free Press. 1988. ISBN: 0-02-914211-3.

Heredia-Langner, Alejandro et. al. Performance of Customized Control Charts to Detect Process Disturbances. John Wiley & Sons: Journal of Quality and Reliability Engineering International. 2001; 17: 205 – 218.

Knudsen, David. *Application of Statistical Quality Control to Improve Yields and Rationalize Testing in a Low Volume Manufacturing Facility.* Boston, MA: MIT Graduate Thesis. 2001.

MacDonald, Francis J. Jr. *An Integrative Approach to Process Parameter Selection in used-Silica Capillary Manufacturing*. Boston, MA: MIT Graduate Thesis. 1992.

Perry, Robert H. and Don W. Green. Perry's Chemical Engineer's Handbook. New York: McGraw Hill. 1997. ISBN: 0-07-049841-5.

Rocke, David M. X_Q and R_Q charts: Robust Control Charts. *The Statistician*, 41, 97-104.

Sall, John et.al. JMP Start Statistics. Pacific Grove, CA.: SAS Institute, 2001.

Welch, Cory. *Application of the Design Structure Matrix: A Case Study in Process Improvement Dynamics*. Boston, MA: MIT Graduate Thesis. 2001.

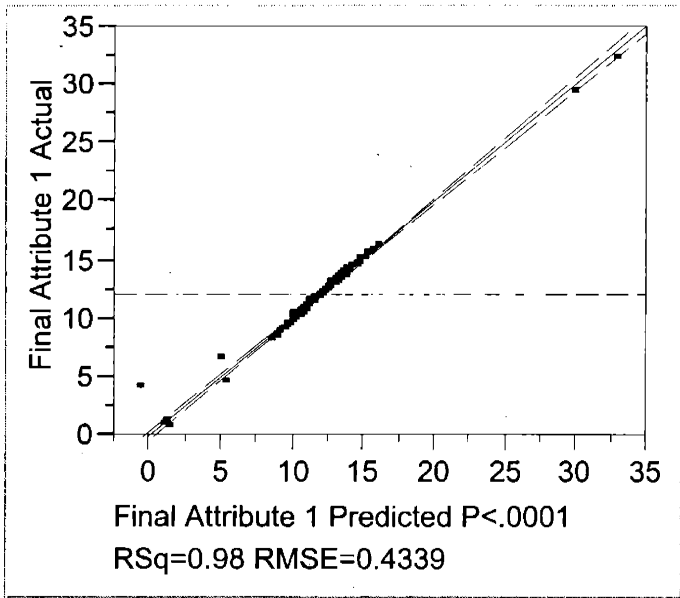
Wise, Stephen A. and Douglas C. Fair. Innovative Control Charting. Milwaukee, WI: ASQ Quality Press, 1998.

Wurl, Robin C., Susan L. Albin, and Irwin J. Shiffer. *Multivariate Monitoring of Batch Process Start-ups.* John Wiley & Sons: Journal of Quality and Reliability Engineering International. 2001; 17: 269-278.

http://www.dsmweb.org/Tutorial - Website for basic DSM information

Vining, G. Geoffrey. Statistical Methods for Engineers. Boston: Duxbury Press. 1998. ISBN: 0-534-23706-1.

Appendix I: Detailed Model Output for Kodak Case Example



Summary of Fit

RSquare	0.983761
RSquare Adj	0.983513
Root Mean Square Error	0.433935
Mean of Response	12.13945
Observations (or Sum Wgts)	200

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	2235.8305	745.277	3957.925
Error	196	36.9068	0.188	Prob > F
C. Total	199	2272.7372		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	193	36.906774	0.191227	.
Pure Error	3	0.000000	0.000000	Prob > F
Total Error	196	36.906774		.
				Max RSq
				1.0000

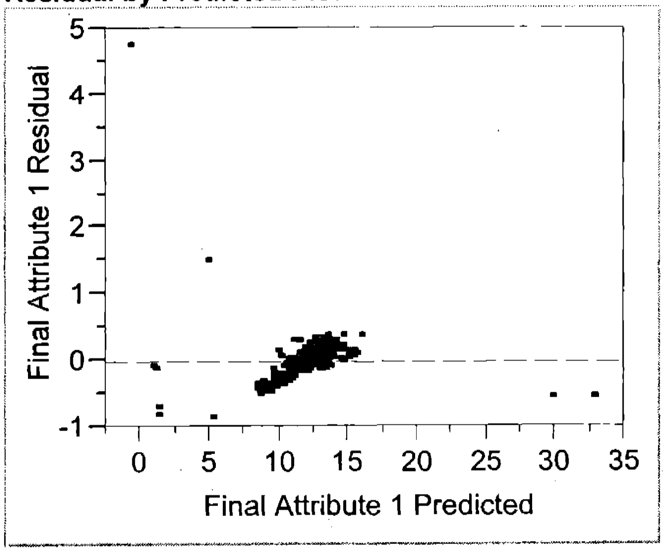
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.0753595	0.107569	10.00	<.0001
Final Attribute 3	0.6048054	0.031546	19.17	<.0001
Final Attribute 4	-0.662893	0.021625	-30.65	<.0001
(Final Attribute 3-7.2448)*(Final Attribute 4+9.9714)	-0.011665	0.001313	-8.88	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 3	1	1	69.21486	367.5779	<.0001
Final Attribute 4	1	1	176.94568	939.7016	<.0001
Final Attribute 3*Final Attribute 4	1	1	14.85945	78.9137	<.0001

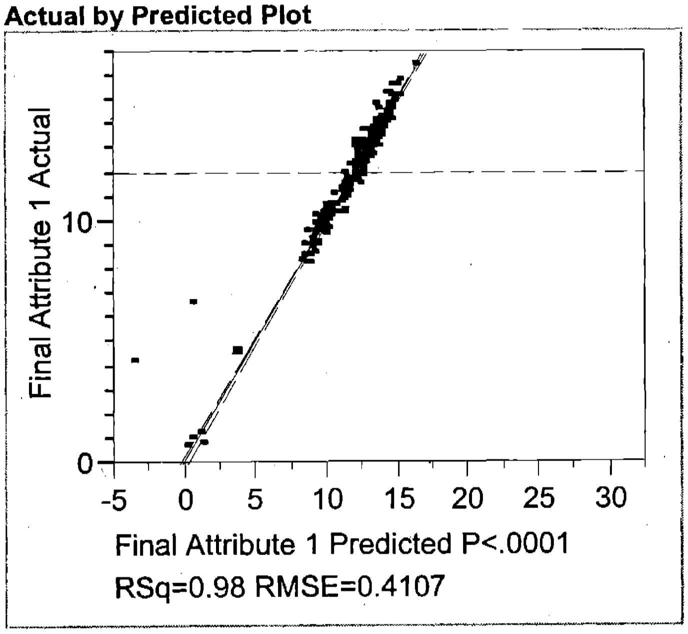
Residual by Predicted Plot



Because residual plot is not random, must remove outliers and change explanatory variables. The resulting model becomes:

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.976829
RSquare Adj	0.976709
Root Mean Square Error	0.41071
Mean of Response	12.05297
Observations (or Sum Wgts)	195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1372.4683	1372.47	8136.382
Error	193	32.5558	0.17	Prob > F
C. Total	194	1405.0241		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	163	27.379744	0.167974	0.9736
Pure Error	30	5.176050	0.172535	Prob > F
Total Error	193	32.555794		0.5636
				Max RSq
				0.9963

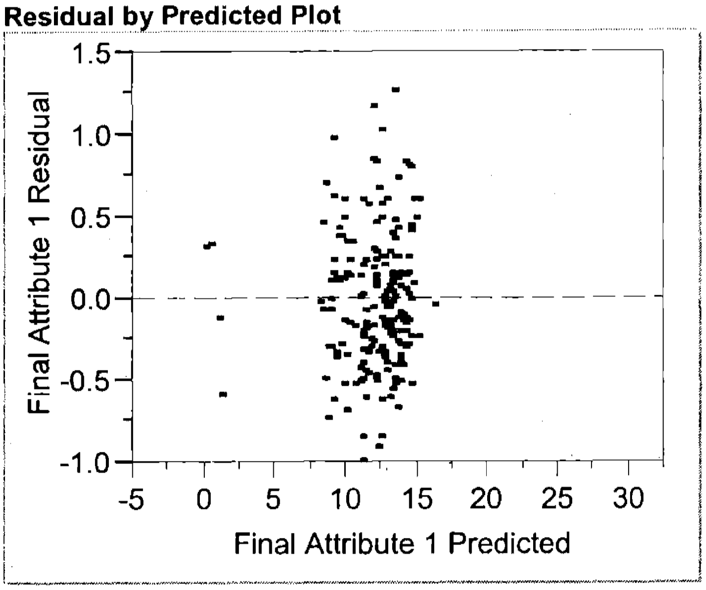
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.2215803	0.123629	9.88	<.0001
Final Attribute 4	-1.085768	0.012037	-90.20	<.0001

Effect Tests

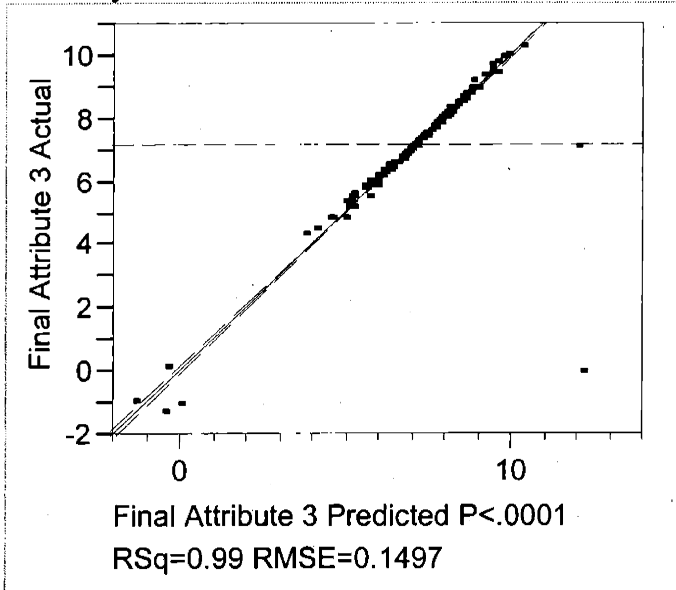
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 4	1	1	1372.4683	8136.382	<.0001

Residual by Predicted Plot



Final Attribute 2 has no correlations within the Final Attribute category.

Response Final Attribute 3
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.993089
RSquare Adj	0.99298
Root Mean Square Error	0.149735
Mean of Response	7.178769
Observations (or Sum Wgts)	195

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	615.33140	205.110	9148.366
Error	191	4.28231	0.022	Prob > F
C. Total	194	619.61370		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	188	4.2823058	0.022778	.
Pure Error	3	0.0000000	0.000000	Prob > F
Total Error	191	4.2823058		.
				Max RSq
				1.0000

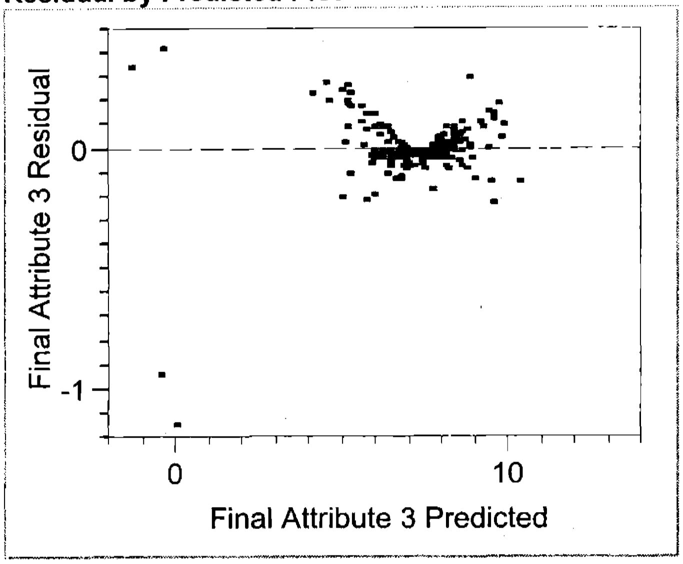
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.136107	0.075467	-1.80	0.0729
Final Attribute 1	1.9395889	0.026389	73.50	<.0001
Final Attribute 4	1.5955887	0.029909	53.35	<.0001
(Final Attribute 1-12.053)*(Final Attribute 4+9.97579)	0.0224724	0.000822	27.34	<.0001

Effect Tests

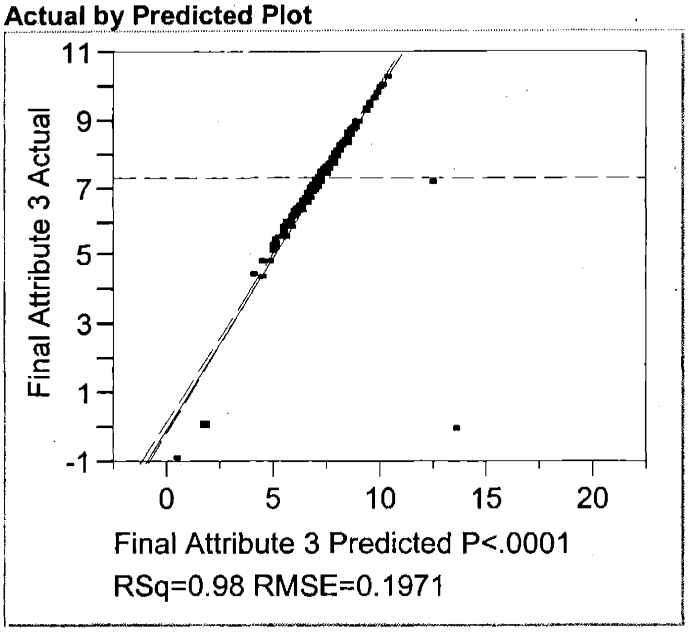
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	121.12142	5402.275	<.0001
Final Attribute 4	1	1	63.80762	2845.957	<.0001
Final Attribute 1*Final Attribute 4	1	1	16.76210	747.6255	<.0001

Residual by Predicted Plot



Again, there is a pattern in the residuals, so the model must be modified:

Response Final Attribute 3
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.980586
RSquare Adj	0.98038
Root Mean Square Error	0.197145
Mean of Response	7.339791
Observations (or Sum Wgts)	191

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	369.06159	184.531	4747.875
Error	188	7.30680	0.039	Prob > F
C. Total	190	376.36839		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	187	7.3068041	0.039074	.
Pure Error	1	0.0000000	0.000000	Prob > F
Total Error	188	7.3068041		.
				Max RSq
				1.0000

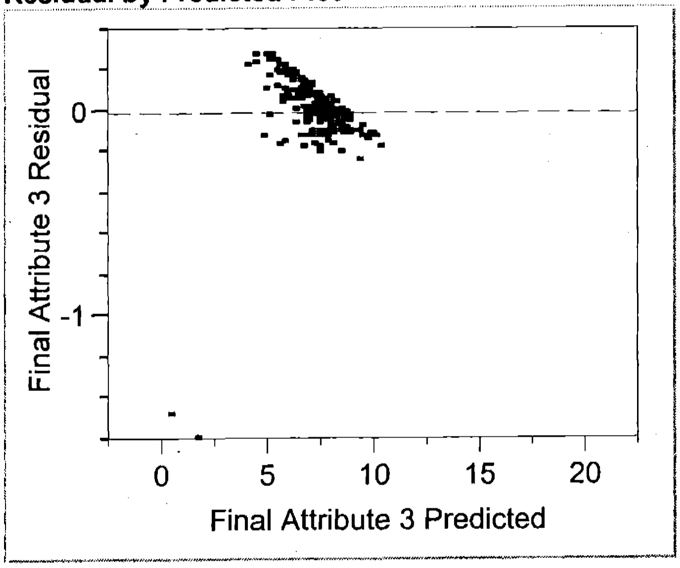
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.765639	0.085121	-8.99	<.0001
Final Attribute 1	1.9155501	0.034788	55.06	<.0001
Final Attribute 4	1.5134542	0.038915	38.89	<.0001

Effect Tests

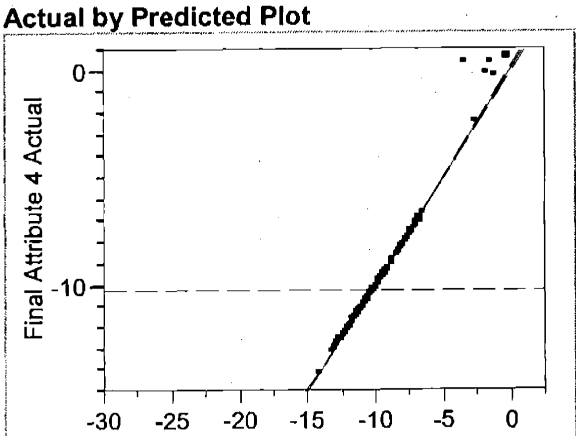
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	117.83953	3031.945	<.0001
Final Attribute 4	1	1	58.78741	1512.567	<.0001

Residual by Predicted Plot



The pattern observed in the residuals is due to a physical limit within the system.

Response Final Attribute 4
Whole Model
Actual by Predicted Plot



Final Attribute 4 Predicted P<.0001
RSq=1.00 RMSE=0.1229

Summary of Fit

RSquare	0.996058
RSquare Adj	0.996017
Root Mean Square Error	0.12285
Mean of Response	-10.1941
Observations (or Sum Wgts)	191

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	717.01170	358.506	23754.41
Error	188	2.83733	0.015	Prob > F
C. Total	190	719.84903		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	187	2.8373297	0.015173	.
Pure Error	1	0.0000000	0.000000	Prob > F
Total Error	188	2.8373297		.
				Max RSq
				1.0000

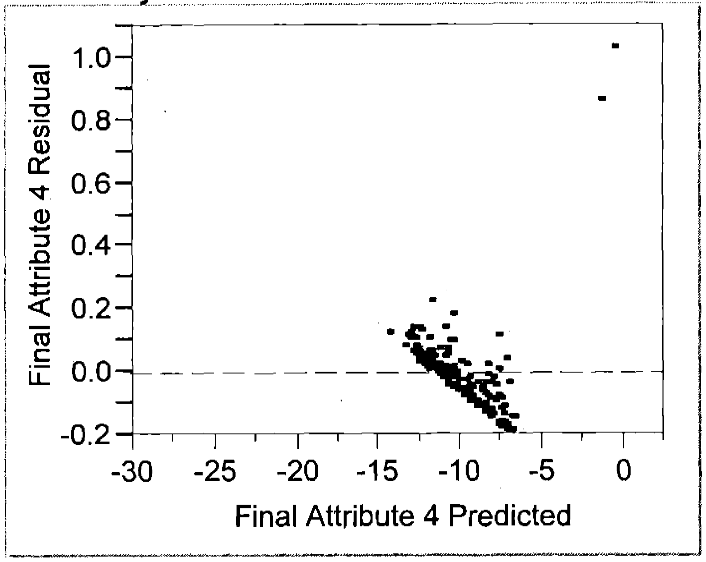
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5153253	0.051106	10.08	<.0001
Final Attribute 1	-1.22281	0.009768	-125.2	<.0001
Final Attribute 3	0.5876945	0.015111	38.89	<.0001

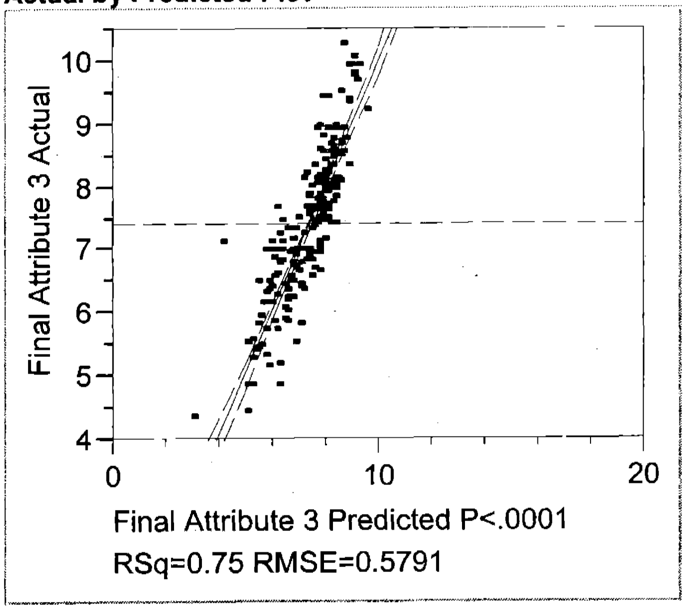
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	236.51982	15671.68	<.0001
Final Attribute 3	1	1	22.82794	1512.567	<.0001

Residual by Predicted Plot



Response Final Attribute 3
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.75458
RSquare Adj	0.753267
Root Mean Square Error	0.579069
Mean of Response	7.421429
Observations (or Sum Wgts)	189

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	192.79535	192.795	574.9581
Error	187	62.70496	0.335	Prob > F
C. Total	188	255.50031		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	155	57.441048	0.370587	2.2528
Pure Error	32	5.263917	0.164497	Prob > F
Total Error	187	62.704965		0.0043
				Max RSq
				0.9794

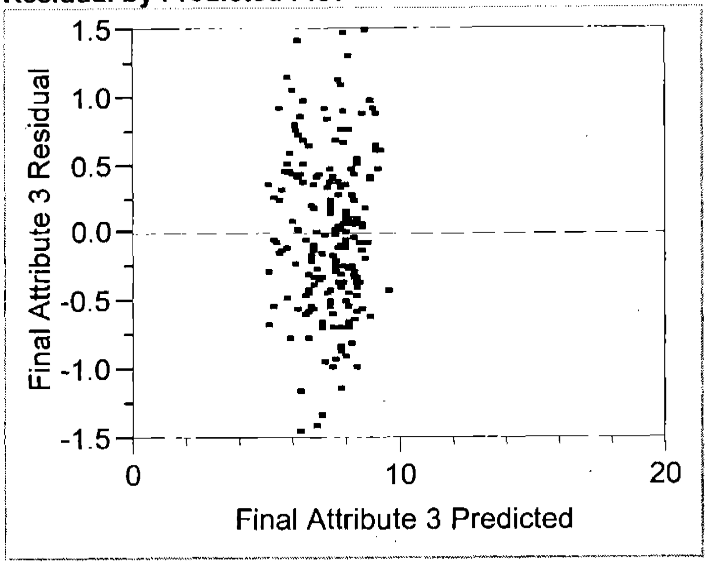
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5813367	0.288355	2.02	0.0452
Final Attribute 1	0.5513027	0.022992	23.98	<.0001

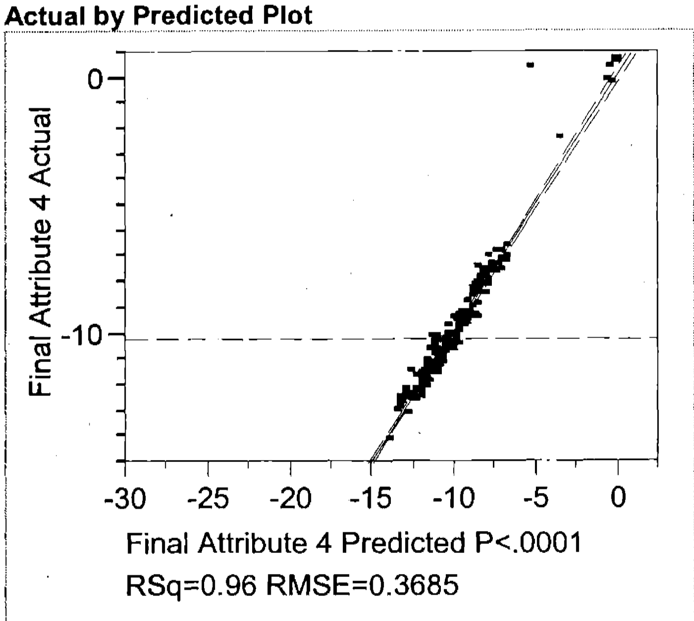
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	192.79535	574.9581	<.0001

Residual by Predicted Plot



Response Final Attribute 4
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.964346
RSquare Adj	0.964158
Root Mean Square Error	0.368504
Mean of Response	-10.1941
Observations (or Sum Wgts)	191

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	694.18377	694.184	5111.996
Error	189	25.66527	0.136	Prob > F
C. Total	190	719.84903		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	157	23.722050	0.151096	2.4882
Pure Error	32	1.943217	0.060726	Prob > F
Total Error	189	25.665267		0.0017
				Max RSq
				0.9973

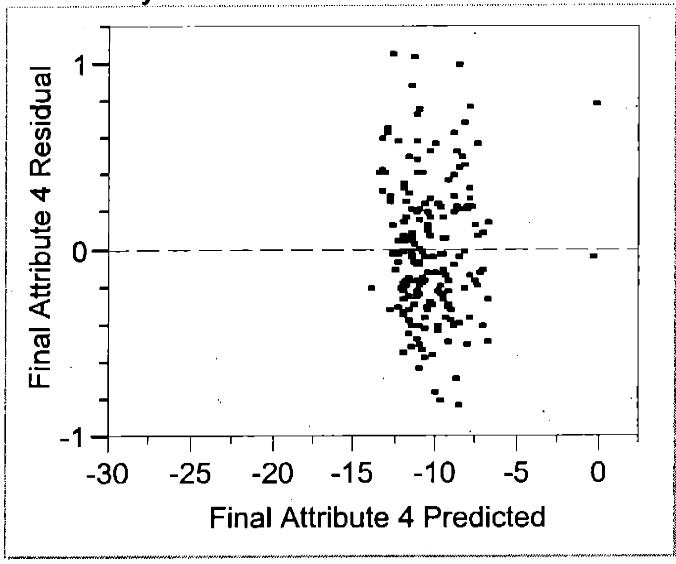
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5912507	0.153187	3.86	0.0002
Final Attribute 1	-0.877885	0.012278	-71.50	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	694.18377	5111.996	<.0001

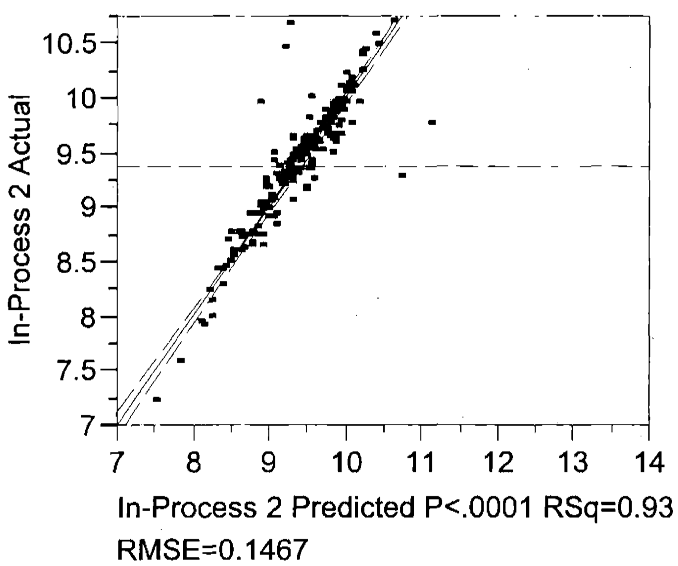
Residual by Predicted Plot



Insert DSM Matrix Analysis to discard half of these variables (pick 1 or 4 and 2)

- See bivariate plot for In-Process 1 relationship

Response In-Process 2
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.92998
RSquare Adj	0.929219
Root Mean Square Error	0.14667
Mean of Response	9.386471
Observations (or Sum Wgts)	187

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	52.571660	26.2858	1221.914
Error	184	3.958210	0.0215	Prob > F
C. Total	186	56.529871		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	183	3.9582103	0.021630	.
Pure Error	1	0.0000000	0.000000	Prob > F
Total Error	184	3.9582103		.
				Max RSq
				1.0000

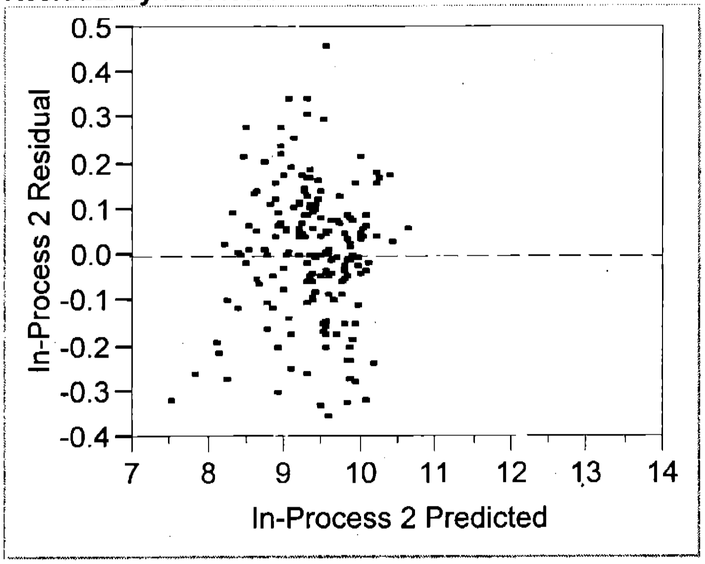
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8513854	0.18976	4.49	<.0001
In-Process 1	0.4241317	0.043939	9.65	<.0001
In-Process 3	0.3737237	0.047105	7.93	<.0001

Effect Tests

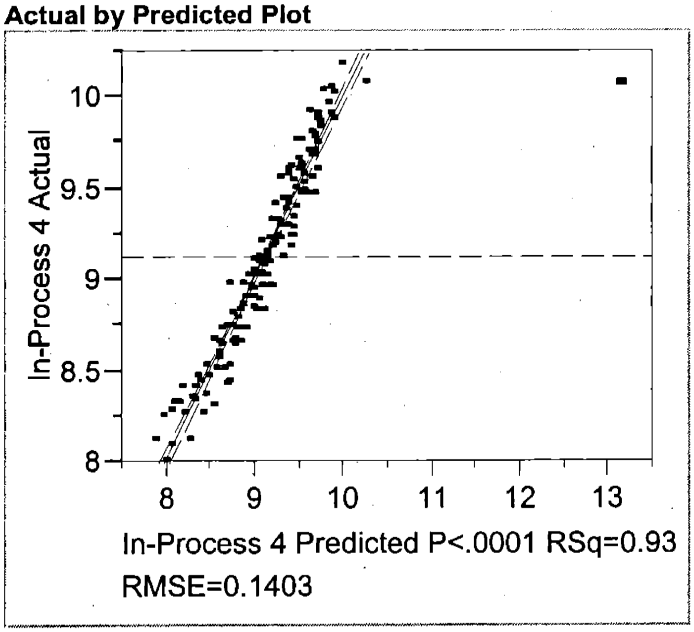
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 1	1	1	2.0043687	93.1744	<.0001
In-Process 3	1	1	1.3541010	62.9463	<.0001

Residual by Predicted Plot



In-Process 3 correlates well with In-Process 2 alone or 1 & 2 together.

Response In-Process 4
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.934691
RSquare Adj	0.933615
Root Mean Square Error	0.140316
Mean of Response	9.123011
Observations (or Sum Wgts)	186

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	51.284381	17.0948	868.2567
Error	182	3.583333	0.0197	Prob > F
C. Total	185	54.867714		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	135	3.5833325	0.026543	.
Pure Error	47	0.0000000	0.000000	Prob > F
Total Error	182	3.5833325		.
				Max RSq
				1.0000

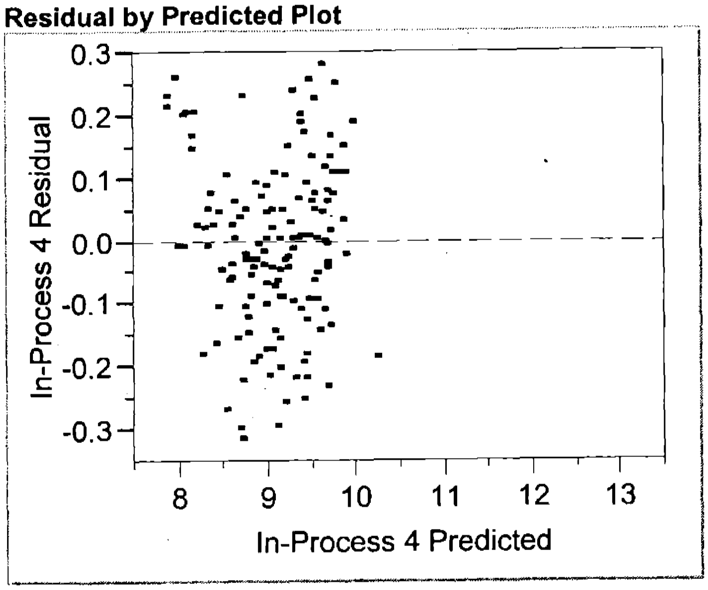
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.0111033	0.16332	6.19	<.0001
In-Process 5	0.7627787	0.037182	20.51	<.0001
In-Process 6	0.1863022	0.023645	7.88	<.0001
(In-Process 5-7.68108)*(In-Process 6-12.544)	-0.197423	0.01038	-19.02	<.0001

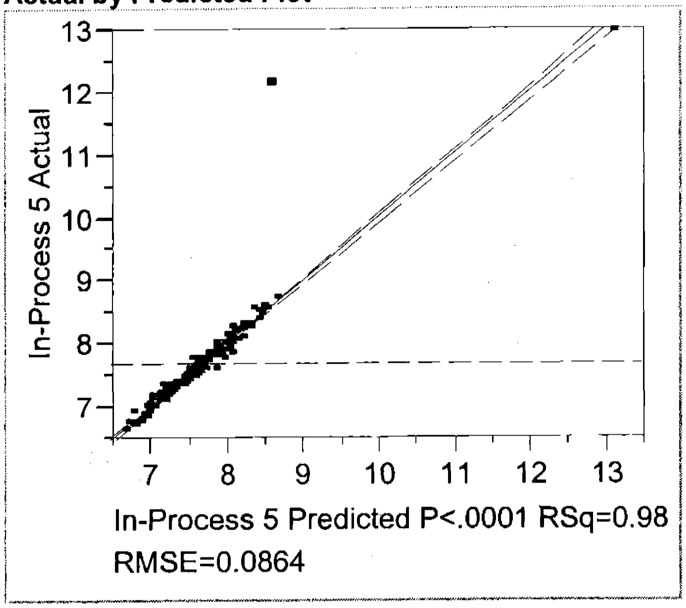
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 5	1	1	8.2860680	420.8553	<.0001
In-Process 6	1	1	1.2223168	62.0823	<.0001
In-Process 5*In-Process 6	1	1	7.1228664	361.7754	<.0001

Residual by Predicted Plot



Response In-Process 5
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.98151
RSquare Adj	0.981205
Root Mean Square Error	0.086364
Mean of Response	7.681075
Observations (or Sum Wgts)	186

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	72.058294	24.0194	3220.307
Error	182	1.357491	0.0075	Prob > F
C. Total	185	73.415785		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	135	1.3574905	0.010055	.
Pure Error	47	0.0000000	0.000000	Prob > F
Total Error	182	1.3574905		.
				Max RSq
				1.0000

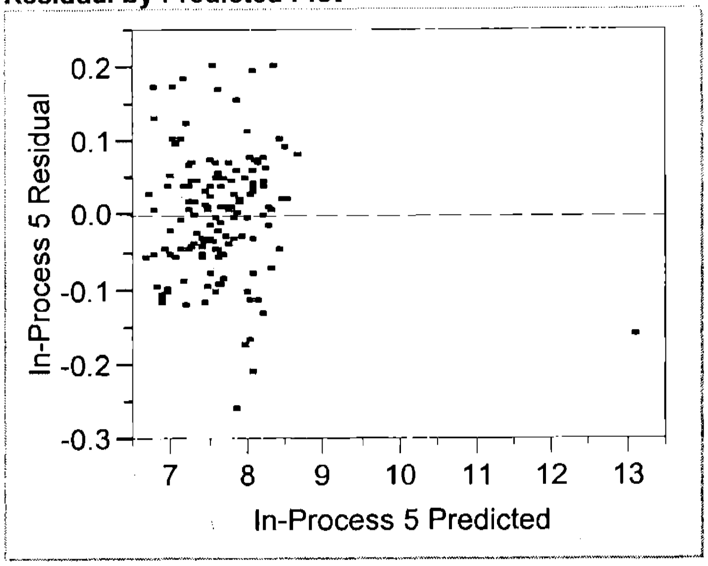
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.193062	0.1083	-1.78	0.0763
In-Process 4	0.9121242	0.024892	36.64	<.0001
In-Process 6	-0.342419	0.019917	-17.19	<.0001
In-Process 7	0.3558204	0.007005	50.80	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 4	1	1	10.015213	1342.749	<.0001
In-Process 6	1	1	2.204682	295.5837	<.0001
In-Process 7	1	1	19.246857	2580.444	<.0001

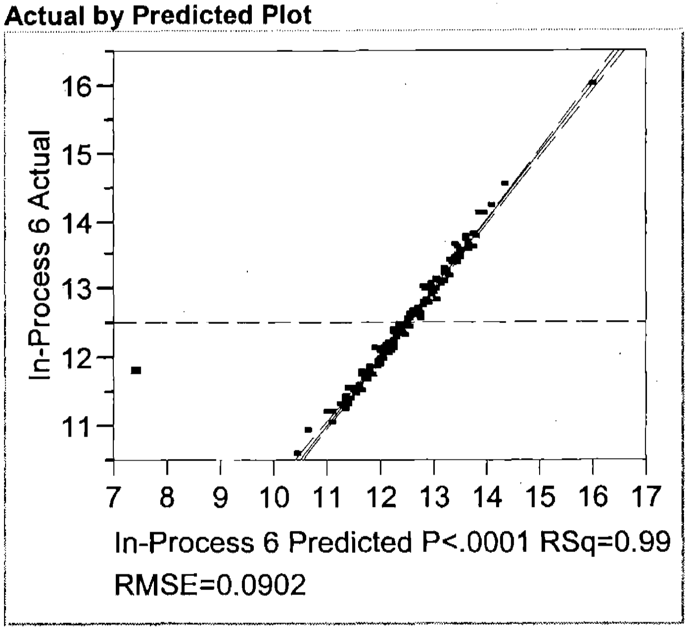
Residual by Predicted Plot



Response In-Process 6

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.9886
RSquare Adj	0.988348
Root Mean Square Error	0.090213
Mean of Response	12.54398
Observations (or Sum Wgts)	186

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	127.73720	31.9343	3923.901
Error	181	1.47305	0.0081	Prob > F
C. Total	185	129.21026		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	134	1.4730517	0.010993	.

Source	DF	Sum of Squares	Mean Square	F Ratio
Pure Error	47	0.0000000	0.000000	Prob > F
Total Error	181	1.4730517		
				Max RSq
				1.0000

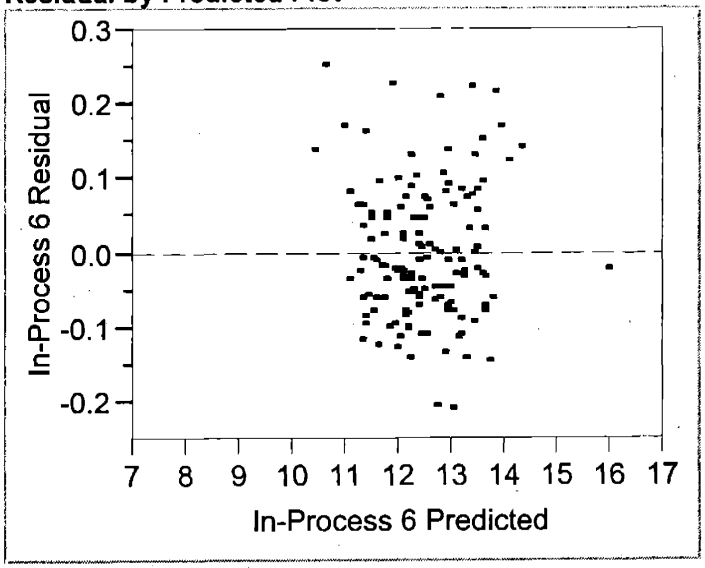
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0658264	0.114187	0.58	0.5650
In-Process 4	1.1983509	0.044108	27.17	<.0001
In-Process 5	-1.316272	0.05128	-25.67	<.0001
In-Process 7	1.0847631	0.019941	54.40	<.0001
(In-Process 5-7.68108)*(In-Process 7-10.8148)	-0.097571	0.003689	-26.45	<.0001

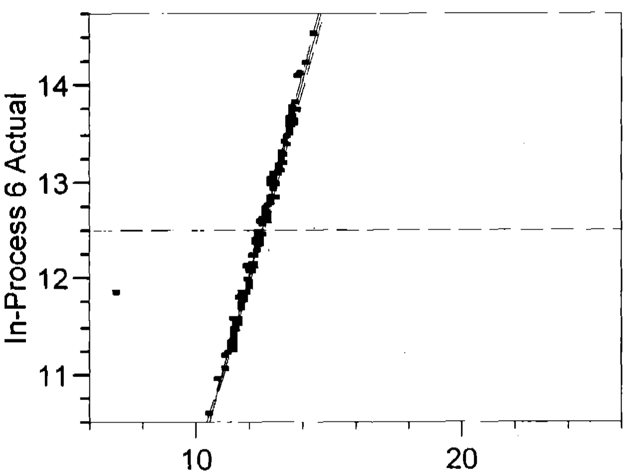
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 4	1	1	6.007170	738.1260	<.0001
In-Process 5	1	1	5.362114	658.8654	<.0001
In-Process 7	1	1	24.082520	2959.12	<.0001
In-Process 5*In-Process 7	1	1	5.692521	699.4638	<.0001

Residual by Predicted Plot



Response In-Process 6
Whole Model
Actual by Predicted Plot



In-Process 6 Predicted P<.0001 RSq=0.99
RMSE=0.0855

Summary of Fit

RSquare	0.988712
RSquare Adj	0.988525
Root Mean Square Error	0.085494
Mean of Response	12.5253
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	115.87865	38.6262	5284.618
Error	181	1.32296	0.0073	Prob > F
C. Total	184	117.20161		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	134	1.3229614	0.009873	.
Pure Error	47	0.0000000	0.000000	Prob > F
Total Error	181	1.3229614		.
				Max RSq
				1.0000

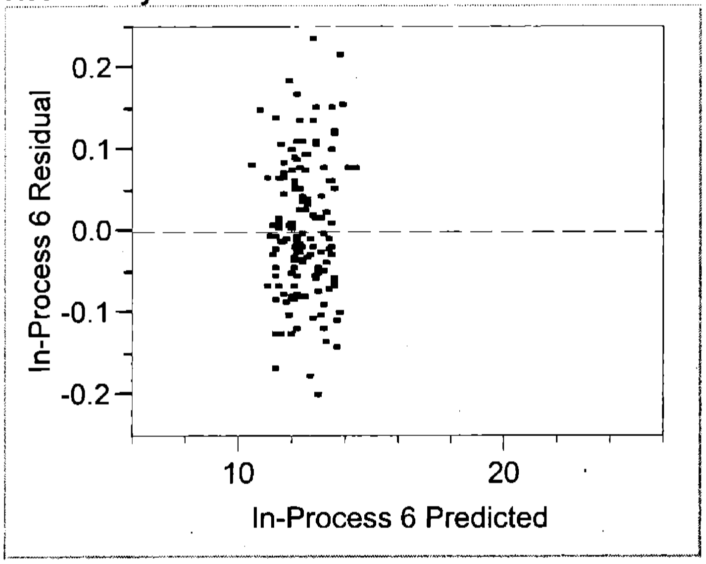
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.068801	0.108128	-0.64	0.5254
In-Process 4	1.2087591	0.041255	29.30	<.0001
In-Process 5	-1.333088	0.048305	-27.60	<.0001
In-Process 7	1.097584	0.019095	57.48	<.0001

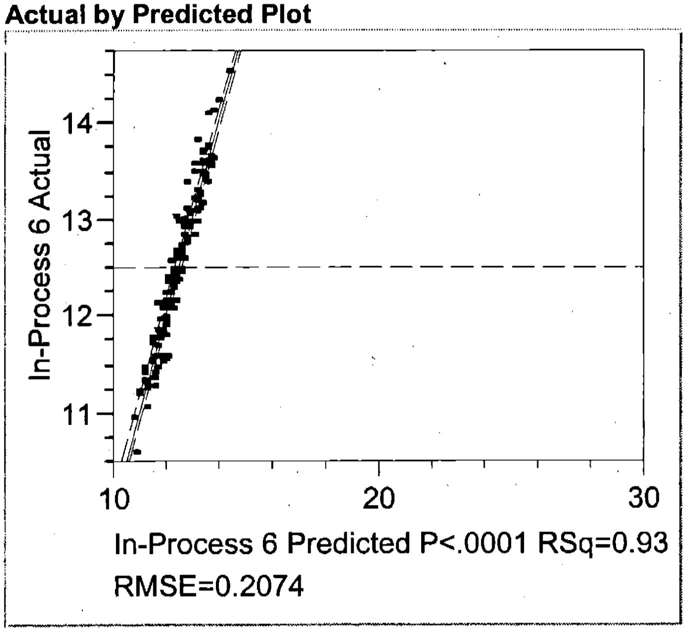
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 4	1	1	6.274694	858.4677	<.0001
In-Process 5	1	1	5.566645	761.5965	<.0001
In-Process 7	1	1	24.150173	3304.088	<.0001

Residual by Predicted Plot



Response In-Process 6
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.932839
RSquare Adj	0.932472
Root Mean Square Error	0.207396
Mean of Response	12.5253
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	109.33023	109.330	2541.794
Error	183	7.87138	0.043	Prob > F
C. Total	184	117.20161		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	100	6.4073902	0.064074	3.6326

Source	DF	Sum of Squares	Mean Square	F Ratio
Pure Error	83	1.4639917	0.017638	Prob > F
Total Error	183	7.8713819		<.0001
				Max RSq
				0.9875

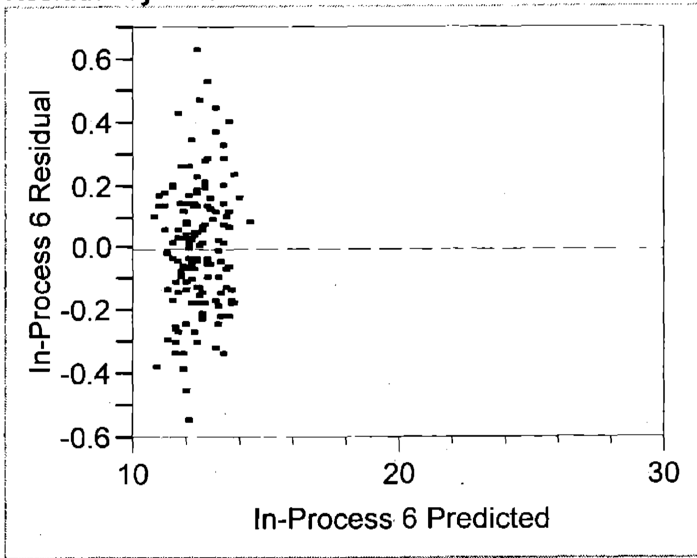
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.0372772	0.228373	4.54	<.0001
In-Process 7	1.0708877	0.021241	50.42	<.0001

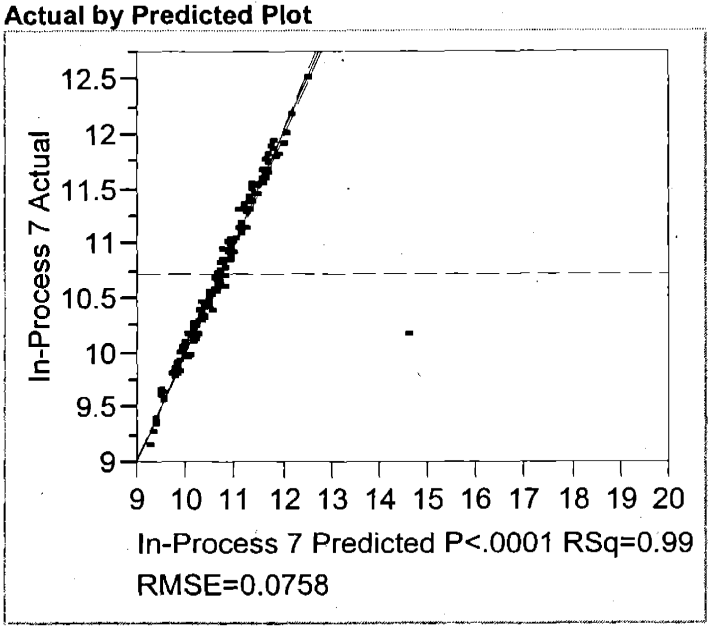
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 7	1	1	109.33023	2541.794	<.0001

Residual by Predicted Plot



Response In-Process 7
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.989079
RSquare Adj	0.988898
Root Mean Square Error	0.075843

Mean of Response	10.72757
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	94.293865	31.4313	5464.259
Error	181	1.041141	0.0058	Prob > F
C. Total	184	95.335005		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	134	1.0411408	0.007770	.
Pure Error	47	0.0000000	0.000000	Prob > F
Total Error	181	1.0411408		.
				Max RSq
				1.0000

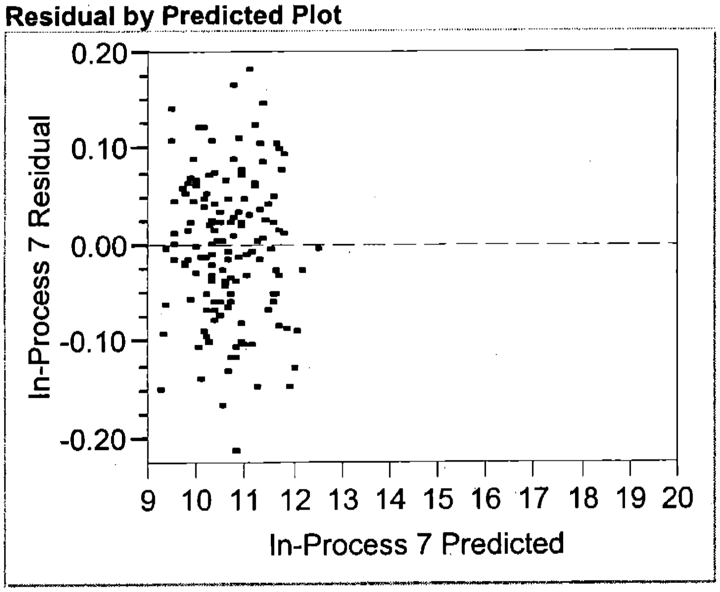
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0889352	0.095802	0.93	0.3545
In-Process 4	-1.028866	0.042939	-23.96	<.0001
In-Process 5	1.2022917	0.039711	30.28	<.0001
In-Process 6	0.8637739	0.015027	57.48	<.0001

Effect Tests

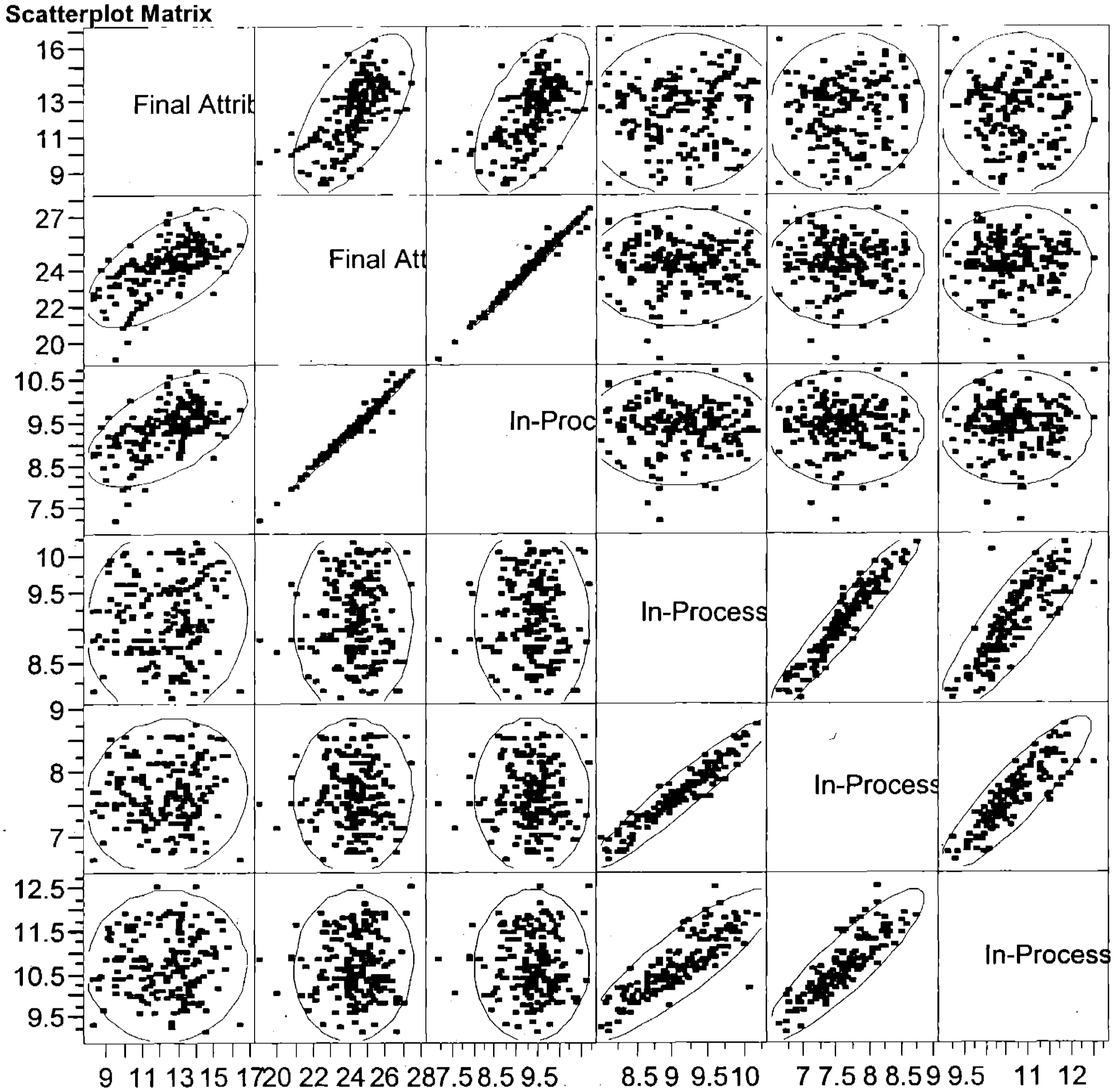
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
In-Process 4	1	1	3.302498	574.1319	<.0001
In-Process 5	1	1	5.272640	916.6366	<.0001
In-Process 6	1	1	19.005642	3304.088	<.0001

Residual by Predicted Plot

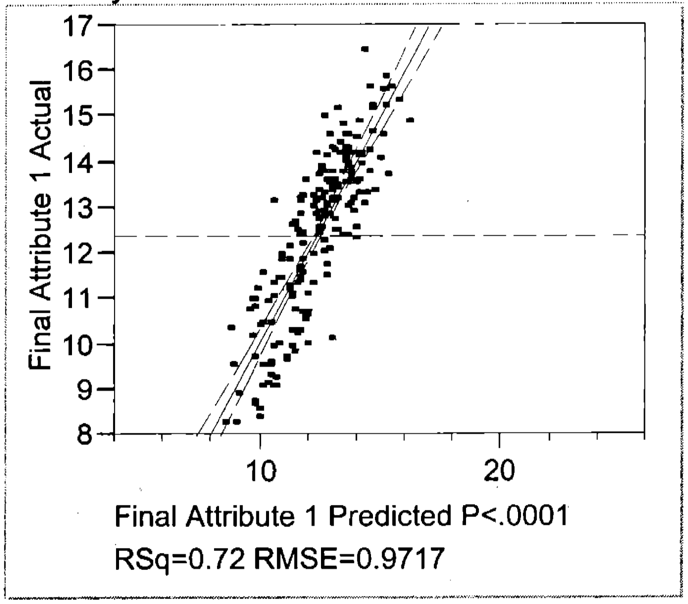


Next level of analysis

Scatterplot Matrix



Response Final Attribute 1
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.724816
RSquare Adj	0.721792
Root Mean Square Error	0.97168
Mean of Response	12.4007
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	452.60789	226.304	239.6878
Error	182	171.83732	0.944	Prob > F
C. Total	184	624.44521		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	180	171.28607	0.951589	3.4525
Pure Error	2	0.55125	0.275625	Prob > F
Total Error	182	171.83732		0.2511
				Max RSq
				0.9991

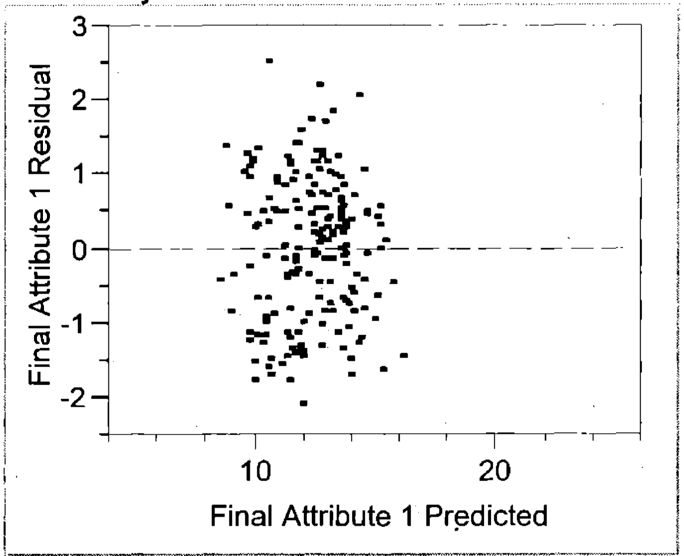
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-18.06136	1.444549	-12.50	<.0001
Final Attribute 2	9.2541439	0.569614	16.25	<.0001
In-Process 2	-20.72068	1.398513	-14.82	<.0001

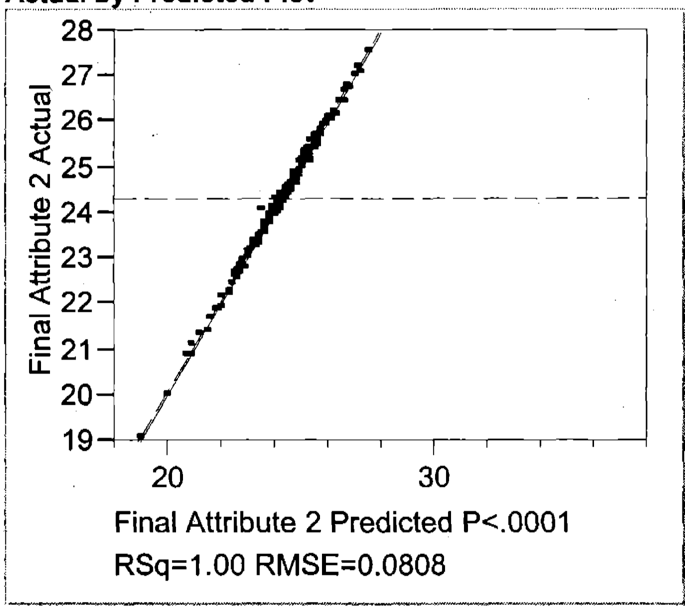
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 2	1	1	249.20502	263.9433	<.0001
In-Process 2	1	1	207.26259	219.5204	<.0001

Residual by Predicted Plot



Response Final Attribute 2
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.996495
RSquare Adj	0.996456
Root Mean Square Error	0.08078
Mean of Response	24.30195
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	337.64828	168.824	25872
Error	182	1.18762	0.007	Prob > F
C. Total	184	338.83590		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	181	1.1876159	0.006561	.
Pure Error	1	0.0000000	0.000000	Prob > F

Source	DF	Sum of Squares	Mean Square	F Ratio
Total Error	182	1.1876159		.
				Max RSq
				1.0000

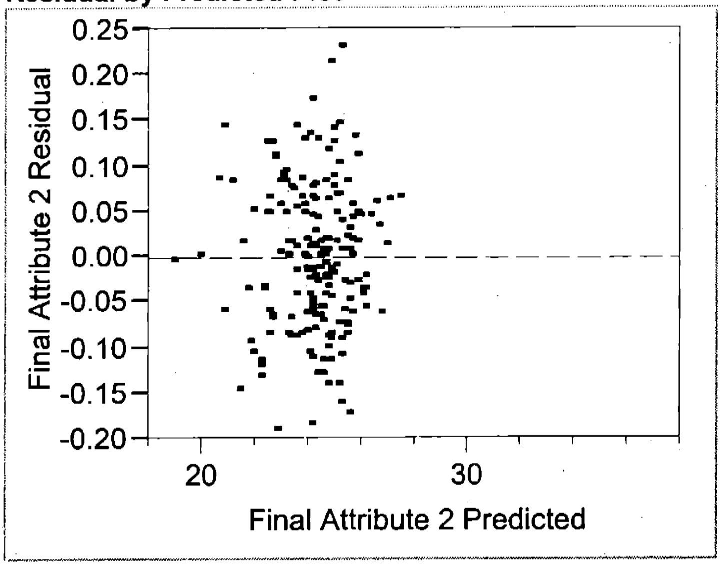
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.7113921	0.103522	16.53	<.0001
Final Attribute 1	0.063958	0.003937	16.25	<.0001
In-Process 2	2.3229631	0.013121	177.04	<.0001

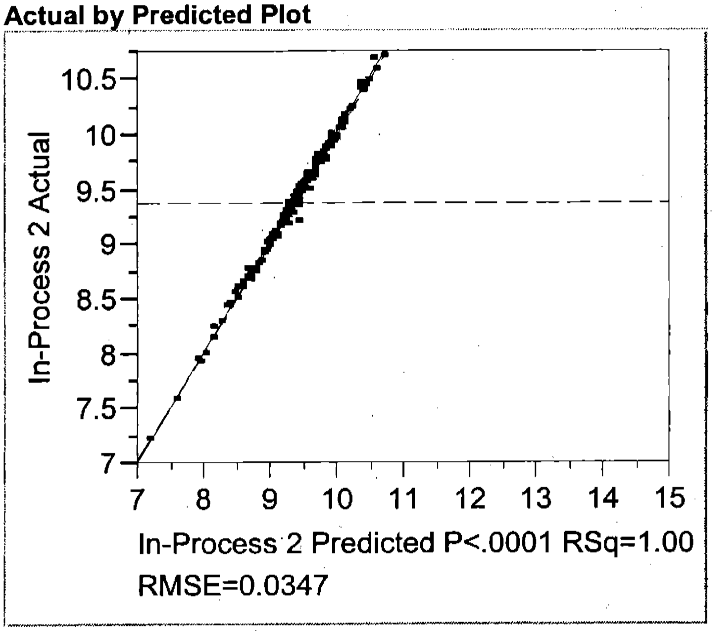
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	1.72233	263.9433	<.0001
In-Process 2	1	1	204.51924	31342.2	<.0001

Residual by Predicted Plot



Response In-Process 2
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.996107
RSquare Adj	0.996064

Root Mean Square Error	0.034674
Mean of Response	9.383459
Observations (or Sum Wgts)	185

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	55.991771	27.9959	23285.67
Error	182	0.218815	0.0012	Prob > F
C. Total	184	56.210586		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	181	0.21881486	0.001209	.
Pure Error	1	0.00000000	0.000000	Prob > F
Total Error	182	0.21881486		.
				Max RSq
				1.0000

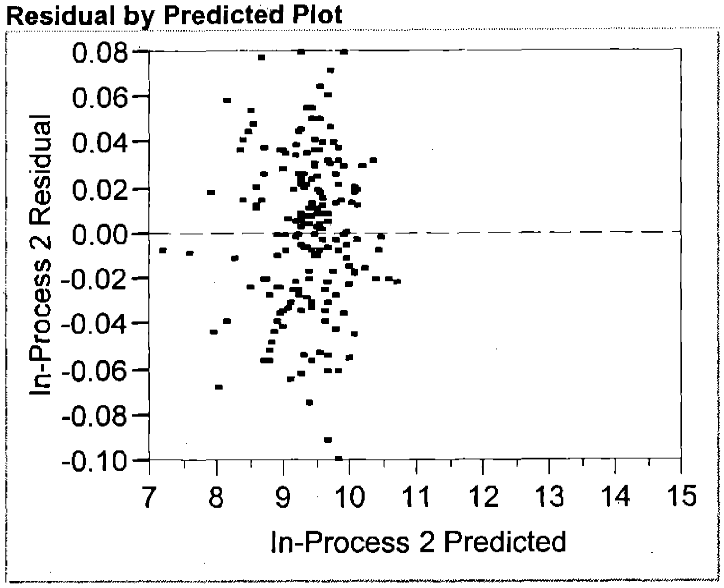
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.69056	0.04816	-14.34	<.0001
Final Attribute 1	-0.026385	0.001781	-14.82	<.0001
Final Attribute 2	0.4279993	0.002418	177.04	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Final Attribute 1	1	1	0.263925	219.5204	<.0001
Final Attribute 2	1	1	37.682088	31342.2	<.0001

Residual by Predicted Plot



The remainder of the In-Process Data is only correlated with other In-Process variables.