Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Improving Quality via Matching: A Case Study Integrating Supplier and Manufacturer Quality Performance

Robert Plante, Herbert Moskowitz, Jen Tang, Jeff Duffy,

Improving Quality via Matching: A Case Study Integrating Supplier and Manufacturer Quality Performance

Robert Plante • Herbert Moskowitz • Jen Tang • Jeff Duffy
Krannert Graduate School of Management, Purdue University, 1310 Krannert Building, West Lafayette, Indiana 47907-1310

The relentless pursuit of increased product quality via continuous improvement is an important long-term strategy for achieving competitive advantage. However, manufacturers must still achieve high product quality in the short run. Hence, short-run quality improvement strategies are necessary, and, if possible, should complement (rather than substitute for) a longer run continuous improvement strategy with suppliers. We propose a novel short-run quality improvement strategy with suppliers which is based on a combinatorial optimization model for determining the optimal matching of raw material batches (from suppliers) with process variable parameters (of the manufacturer) by exploiting their interaction effects. Although the model can be difficult to solve, we develop sufficient conditions for a special case where the attainment of an optimal solution is straightforward. We demonstrate the use of the proposed approach in an actual pharmaceutical process. Using data from the pharmaceutical case, we develop several simulation scenarios where matching is shown (predicted) to greatly improve process yield. We also discuss the use of matching to illuminate various strategies for long-term continuous improvement of supplier and manufacturer processes in general.
(*Matching*; *Quality Improvement*; *Combinatorial Optimization*; *Heuristics*; *Logit Model*; *Likelihood Ratio*)

1. Introduction

For many manufacturers, an important component of a successful continuous quality improvement program is the receipt of high quality incoming raw materials that are important performance-influencing inputs for products/processes. One facet of a manufacturer's cost that is significantly affected by the quality of incoming raw materials is scrap and rework, a large percentage of which can be directly attributed to raw material variation (Kelly 1993). For example, firms such as Ford and Xerox have spent billions of dollars each year to inspect and rework shoddy components (Gabor 1989). Since the quality and production cycle time of their end products are greatly affected by the quality of the raw materials they receive from their suppliers (Marsh and Tucker 1990), manufacturers work closely with suppliers to continuously improve the quality of incoming raw material and also aid suppliers in fostering and maintaining continuous improvement efforts through supplier certification programs and those such as Motorola's training of suppliers in statistical process control methods (Kelly 1994).

Such integrated training programs aid suppliers in the long-term continuous improvement of their processes. Ultimately, such improvements provide high-quality resources that are not only important in increasing the overall quality to supplier and manufacturer, but also additional benefits, such as

1523-4614/99/0101/0036$05.00

decreased product cycle times. Although such programs have a track record of proven success, they do not, necessarily, address the issue of achieving immediate yield improvements. Nor do they adequately address the issue of variance reduction in resource yielding processes that are unavoidably subject to the vagaries of the natural environment within which they operate, such as the variation in the quality of rubber tree resin, an essential ingredient in producing rubber, which, in turn, is an important component of many products.

We develop a methodology for improving the quality of a manufacturer's product by matching variations in raw material quality provided by suppliers. Namely, rather than randomly selecting raw material batches from inventory to produce a production batch, raw material batches are consciously selected such that the expected quality of product is maximized over the horizon of raw material inventory (number of batches of raw material held in inventory).

We first develop a general combinatorial optimization model that can be used to improve product quality by matching raw materials when it is not technologically nor economically feasible for a supplier, in the short run, to reduce the variation of raw material output by attacking root causes.

We use two scenarios to illustrate where matching is appropriate. The first discusses the criterion of minimum expected cost, where cost is modeled via a quadratic loss function (Taguchi 1987). We then consider the scenario where process yield is modeled as a second-order polynomial function of the design parameters (raw material quality characteristics and process parameters), and develop a model for determining the raw material matching that maximizes yield. A case study that considers the maximization of the probability of attaining a high yield production run for an actual pharmaceutical product is used to illustrate this formulation and solution approach.

We next consider the effect on process yield when supplier processes are controlled and when process control settings are *adjusted* to match the known values of incoming raw material quality characteristics from suppliers. This leads to a strategy for assessing the impact of future supplier process improvement on process yield. This also provides a mechanism for identifying supplier collaboration opportunities for improving the quality of a manufacturer's product.

2. Problem Formulation

Consider a batch manufacturing process where a manufactured product has a single quality characteristic measure denoted by Y, that is influenced by the quality characteristics of a set of M raw materials, and a set of N-M process parameters. The single quality characteristic of the ith raw material is denoted by X_i with mean μ_i and variance σ_i^2, $i = 1, \ldots, M$. The nth process parameter is denoted X_{M+n}, with mean μ_{M+n} and variance σ^2_{M+n}, $n = 1, \ldots, N\text{-}M$. Supplier raw material quality characteristics are plausibly assumed to be independent, and are called the (explanatory) *variables* of Y.

One batch of *each* raw material is required to produce one production batch. There is a total of T batches of each raw material for T production runs. For the kth batch ($k = 1, \ldots, T$) of the ith raw material ($i = 1, \ldots, M$), estimates of the mean and variance of the associated quality characteristic are available and are denoted by μ_{ik} and σ^2_{ik}, respectively. These values could be the mean and variance of the supplier's process when the batch was provided or could be obtained by sampling from raw materials provided by the supplier.

The estimated mean and variance of the value for raw material i used in the tth production batch are denoted by μ'_{it} and σ'^2_{it}, respectively, for $i = 1, 2, \ldots, M$. We denote the *expected* quality of the tth production batch as $Q(Y_t)$, where $Q(Y_t)$ is in turn a function of the moments of raw material inputs for that batch, including, for example, μ'_{it}, σ'^2_{it}, etc., $i = 1, 2, \ldots, M$.

The matching problem is then to determine the matching of raw material batches that optimizes the expected quality of T production runs. Formally, this problem can be stated as the following constrained optimization model:

$$\text{Max } E\left[\sum_{t=1}^{T} Q(Y_t)\right] \tag{1a}$$

subject to

$$\mu'_{it} = \sum_{k=1}^{T} \delta_{ikt}\mu_{ik} \quad i = 1, \ldots, M,\ t = 1, \ldots, T, \tag{1b}$$

$$\sigma_{it}'^2 = \sum_{k=1}^{T} \delta_{ikt}\sigma_{ik}^2 \quad i = 1, \ldots, M,\ t = 1, \ldots, T, \tag{1c}$$

$$\sum_{t=1}^{T} \delta_{ikt} = 1 \qquad i = 1, \ldots, M,\ k = 1, \ldots, T, \tag{1d}$$

$$\sum_{k=1}^{T} \delta_{ikt} = 1 \qquad i = 1, \ldots, M,\ t = 1, \ldots, T, \tag{1e}$$

where $\delta_{ikt} \in \{0,1\}$ such that, $\delta_{ikt} = 1$ if the kth value of variable i is used in the tth production run, and is 0 otherwise, and $E[\cdot] =$ Expected value.

2.1. Two Examples For *Q(Y)* For *t*th Production Batch

In practice, the exact relationship (functional form) between Y and the explanatory variables X_i's and the corresponding quality function $Q(Y)$ are generally unknown and hence need to be approximated/estimated on a case-to-case basis. Since Y and X's are random variables, their functional forms can be estimated by statistical methods such as regression analysis and response surface methods (Myers and Montgomery 1995). On the other hand, depending on the type of characteristic of Y, $Q(Y)$ can be defined in terms of, for example, expected yield, percent defectives, or loss/cost. The following two examples illustrate ways of obtaining these functions in two commonly encountered scenarios (Askin and Goldberg 1988, Pignatiello 1993, Vining and Myers 1990).

Example 1. Assume a standard multiple linear regression model, namely $Y = \Sigma_{i=1}^{M} a_i X_i + \epsilon$, where ϵ is the error term with mean 0 and known variance σ_ϵ^2. Given a target or nominal value, say τ, of Y, we express the quality (cost) of Y, using a Taguchi (1987) quadratic loss function as

$$Q(Y) = L(Y - \tau)^2, \tag{2}$$

where $Q(Y) =$ cost of quality of Y, and $L =$ quadratic cost coefficient.

In fact, the true loss function of Y is difficult to obtain, but Equation (2) is a truncated Taylor series expansion around τ for the true loss function that can be used as an approximation. Plausibly assuming that the loss function reaches its minimum at the target (hence its first derivative at τ is zero), and by ignoring the third and higher order terms in the series expansion we obtain the quadratic loss function in Equation (2). One clear advantage of using cost/loss for $Q(Y)$ is that it directly demonstrates the effectiveness of matching in minimizing cost. The expected quality loss is then

$$\begin{aligned} E(Q(Y)) &= L\ [\sigma_Y^2 + (\mu_Y - \tau)^2] \\ &= L\left[\sum_{i=1}^{M} a_i^2\ \sigma_i^2 + \sigma_\varepsilon^2 + \left(\sum_{i=1}^{M} a_i\mu_i - \tau\right)^2\right] \\ &= L\left[\sum_{i=1}^{M} a_i^2\ \sigma_i^2 + \sigma_\varepsilon^2 + \left(\sum_{i=1}^{M} a_i^2\ \mu_i^2 + 2\sum_{i=1}^{M}\right.\right. \\ &\quad \left.\left.\cdot \sum_{j>i}^{M} a_i a_j \mu_i \mu_j - 2\tau \sum_{i=1}^{M} a_i\mu_i + \tau^2\right)\right], \end{aligned} \tag{3}$$

where μ_i is the mean of the ith raw material quality characteristic.

Since only cross-product terms in Equation (3) involving the means of each pair of variables, $\mu_i\mu_j$ $(i \neq j)$, are affected by matching, the objective function, $E[\Sigma_{t=1}^{T} Q(Y_t)]$, in Equation (1) can be reduced to

$$E\left[\sum_{t=1}^{T} Q(Y_t)\right] = \sum_{i=1}^{M}\sum_{j>i}^{M} a_i a_j \sum_{k=1}^{T}\sum_{l=1}^{T}\sum_{t=1}^{T} \mu_{ik}\delta_{ikt}\mu_{jl}\delta_{jlt}, \tag{4a}$$

$$= \sum_{i=1}^{M}\sum_{j>i}^{M} a_i a_j \sum_{k=1}^{T}\sum_{l=1}^{T} \mu_{ik}\mu_{jl} \sum_{t=1}^{T} \delta_{ikt}\delta_{jlt}. \tag{4b}$$

For this objective function, an alternative and a more concise formulation for Equation (1) can be obtained by defining $\delta_{ikjl} = \Sigma_{t=1}^{T} \delta_{ikt}\delta_{jlt}$, such that $\delta_{ikjl} = 1$ if and only if $\delta_{jlt} = \delta_{ikt} = 1$ for exactly one t (i.e., production run), and where the kth value of variable i is matched with the lth value of variable j in that production run. Otherwise $\delta_{ikjl} = 0$. Thus, the value of the jth variable that is used to match with the kth mean value of the ith variable, μ_{ik}, is obtained from $\Sigma_{l=1}^{T} \mu_{jl}\delta_{ikjl}$. Consequently, given μ_{ik}'s and σ_{ik}^2's, Y_t is a function of δ_{ikjl}'s and Equation (1) can be reformulated as

$$\text{Min } E\left[\sum_{t=1}^{T} Q(Y_t(\{\delta_{ikjl}\}))\right] = \sum_{i=1}^{M}\sum_{j>i}^{M} a_i a_j \sum_{k=1}^{T}\sum_{l=1}^{T} \mu_{ik}\mu_{jl}\delta_{ikjl} \tag{5a}$$

subject to

$$\sum_{k=1}^{T} \delta_{ikjl} = 1, \quad (j > i,\ i = 1, 2, \ldots, M),\ l = 1, 2, \ldots, T, \tag{5b}$$

$$\sum_{l=1}^{T} \delta_{ikjl} = 1, \quad (j > i,\ i = 1, 2, \ldots, M),\ k = 1, 2, \ldots, T, \tag{5c}$$

$$\delta_{ikjl} + \delta_{ikrs} \leq 1 + \delta_{jlrs}, \quad (i < j < r,\ \{i, j, r\} \subseteq \{1, \ldots, M\}),\ k, l, s = 1, 2, \ldots, T, \tag{5d}$$

$$\delta_{ikjl} \in \{0,1\}, \quad (i, j, j > i, i = 1, 2, \ldots, M),\ k, l = 1, 2, \ldots, T. \tag{5e}$$

Note that Equations (5b) and (5c) are obtained from the definition of δ_{ikjl}, Equations (1d) and (1e). The constraints in Equations (1d)–(1e) and (5b)–(5e) basically enforce matching.

Example 2. The quality characteristic, Y, of the production batch may be influenced by the explanatory variables X_i's through the following function:

$$h(X_1, X_2, \ldots, X_M) = a_0 + \sum_{i=1}^{M} a_i X_i + \sum_{i=1}^{M} \sum_{j>i}^{M} a_{ij} X_i X_j, \tag{6}$$

where the second term on the right side represents the main effects, and the third term represents the second-order interaction effects of the explanatory variables. The expected value of Equation (6) for a production run is

$$E[h(X_1, X_2, \ldots, X_M)] = a_0 + \sum_{i=1}^{M} a_i \mu_i + \sum_{i=1}^{M} \sum_{j>i}^{M} a_{ij} \mu_i \mu_j. \tag{7}$$

Again, only the last term of Equation (7) is affected by matching, hence the objective function in this case is the same as in Equation (5) with $a_i a_j$ replaced by a_{ij}. A possible relationship between $Q(Y)$ and h will be described in the case study in §3.

The model in Equation (1) and hence Equation (5) can be formulated as a multidimensional matching problem (Nemhauser and Wolsey 1988) with $\binom{M}{2}2T$ nodes and $\binom{M}{2}T^2$ edges, resulting in 3,744 nodes and 22,464 edges when producing, for example, $T = 12$ batches with $M = 13$ matching variables. Since there are T values for each of M variables (and hence T production batches) in this matching problem, the number of combinations for the variable values that would have to be examined to find the optimal solution is $(T!)^{M-1}$, which grows exponentially as the size of the problem increases. In general Equation (1) is a combinatorial optimization problem that belongs to the class of NP-complete problems (Papadimitriou and Steiglitz 1982), except for the case where $M = 2$. However, exploiting special properties of the objective function in Equation (5), the size of the matching problem is reduced from $\binom{M}{2}2T$ to M nodes and from $\binom{M}{2}T^2$ to $\binom{M}{2}$ edges. Thus for $T = 12$ batches with $M = 13$ matching variables, the number of nodes and edges to consider are 13 and 78, respectively. Further, under certain conditions, an optimal matching for Equation (5) can be obtained in a straightforward manner. We address these issues as part of the case study presented in the next section.

3. Maximizing Process Yield: A Case Study

We now demonstrate how the matching models in Equations (1) and (7) were used to provide guidance for improving the product quality of a major pharmaceutical manufacturer. The 1993 net sales of this company was \$6.5 billion with a global presence in North America, Japan, Europe, The Middle East, Africa, Asia, and Latin America. Its critical capabilities included scientific innovation, disease prevention and management, and biotechnology expertise. The targeted disease categories for this company were the central nervous system and related diseases, endocrine diseases (including diabetes and osteoporosis), infectious diseases, cancer, and cardiovascular diseases.

In 1992, the patent on an oral antibiotic, which was often prescribed by pediatricians for child related infections (i.e., ear infections), was about to run out. Generic drugs were poised to enter this lucrative market. It was therefore prudent and timely to consider process improvements for manufacturing this drug, resulting in lower costs and higher productivity (higher process yield). It was estimated that a 1% increase in process yield would provide a \$1 million increase in

Figure 1 **Diagram of Critical Steps in Manufacturing Process**

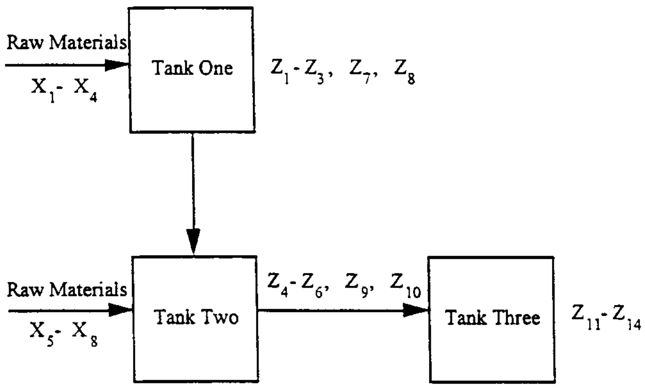


Table 1 **Standard Deviations for Controllable and Uncontrollable Parameters***

Variable	Standard Deviation	Variable	Standard Deviation
X_1	0.0385	Z_1	1.1523
X_2	1.1739	Z_2	447.3765
X_3	1.0143	Z_3	0.2318
X_4	0.8942	Z_4	0.7315
X_5	0.1134	Z_5	148.7407
X_6	0.1010	Z_6	4.8504
X_7	0.9235	Z_7	1.5011
X_8	0.1761	Z_8	1.7081
		Z_9	1.2650
		Z_{10}	2.2048
		Z_{11}	27.8827
		Z_{12}	11.6611
		Z_{13}	2.3383
		Z_{14}	0.1783

*All variables were scaled so that the mean equals zero to protect privacy of the manufacturer.

return. Current process yields were then running between 75% and 95%.

3.1. A Pharmaceutical Batch Process

In this process an oral antibiotic is produced in batches on a dedicated line at a rate of approximately three batches per day. The critical steps in the manufacturing process consist of the chemical processes that are performed in the first three tanks (Figure 1). Four raw materials (with quality characteristics $X_1 - X_4$, respectively) are added to the first tank. Following first tank processing, the batch is then sent to a second tank where further processing is completed, and four more raw materials (with $X_5 - X_8$, respectively) are added. The batch then goes to tank three where further processing occurs. Additional steps follow tank three, but these three operations that take place at the beginning of the manufacturing process are considered to be the most critical by process specialists. At the end of the manufacturing process, the concentration of the production batch is measured, and the yield is calculated.

Data for this example was obtained from a company historical database that consisted of 262 production batches. For each production batch, measurements were available for eight raw materials with quality characteristics $X_1 - X_8$, respectively, and 14 process parameters $X_9 - X_{22}$, which are now denoted by $Z_1 - Z_{14}$ in this section for clear distinction. Although measurable, the manufacturer's process parameters are currently considered uncontrollable or too costly to control. (In §3.5, as part of a simulation study, we will assess the potential impact for controlling some of these process parameters.) Summary statistics for each of the $M = 8$ raw material quality characteristics and $N - M = 14$ process parameters are presented in Table 1.

This data was originally analyzed by company engineers, using multiple regression estimation procedures. Their initial conclusions were that, within the current range of incoming raw material levels (values), there was no significant impact of the quality of raw materials on process yield. Thus, and in keeping with prevailing beliefs, there was no concern about which batch of a raw material was mixed with other batches, since there were no observed or expected effects of such mixing on process yield. Following this analysis, the authors were approached by company representatives and asked to study the same process data and provide analysis and recommendations, which are described below.

3.2. Estimating a Model of Process Yield

Before the optimization model in Equation (5) can be implemented, a model for process yield must be estimated to predict the yield of a production batch with specified values of the raw material quality characteristics and process parameters. Prior analysis had revealed a large amount of process variation, which created difficulty in discerning variable effects on process yield; consequently yield was modeled as a discrete variable in this study. The median of the yield distribution was used to partition the production batches into two groups of equal population, which we denote as high yield and low yield. If yield was greater than 85.45%, then the batch was classified as high yield, otherwise it was classified as low yield. Alternatively, by removing observations, one could partition batches into the 30 highest yield batches and the 30 lowest yield batches, and perform the analysis on this reduced, but presumably more discriminating observation set.

The quality of a production batch was then defined as follows:

$$Y = \begin{cases} 1 \text{ if high yield,} \\ 0 \text{ if low yield.} \end{cases} \tag{8}$$

Since yield is now represented as a binary variable, specialized estimation procedures such as two group discriminant, LOGIT and PROBIT analyses may be used. In the next section we describe the use of the LOGIT procedure for estimating the likelihood function of the expected yield in terms of the raw material quality characteristics and process parameters.

3.3. Logistic Model to Predict Expected Yield

The expected yield, $E(Y)$, of Y in Equation (8) is interpreted as the probability (likelihood) that a production batch belongs to the high yield category. Since $0 \leq E(Y) \leq 1$, the following nonlinear multiple logistic model (Neter et al. 1990) was employed to predict whether a production batch, say the tth, will have high or low yield:

$$E(Y_t) = \frac{1}{1 + \exp\{h(X_{1t}, X_{2t}, \ldots, X_{Mt})\}}, \tag{9}$$

where $h(X_{1t}, X_{2t}, \ldots, X_{Mt})$ is defined in Equation (6). The tth production batch is predicted to belong to the high yield group if the predicted $E(Y_t) > 0.5$, and to the low yield group if the predicted $E(Y_t) \leq 0.5$.

The significant factors and their coefficient estimates in the logistic model were determined by using stepwise discrimination. After performing this procedure, the initial set of 254 terms in Equation (6) was reduced to 47. Of these significant terms, 28 included a raw material quality characteristic term. Estimated coefficients, $\hat{a}_i$ and $\hat{a}_{ij}$, of the significant factors in the logit model are presented in Table 2. Re-substituting each observation into estimated Equation (6), and using the prediction in Equation (9) to classify each production batch, only 3.4% of the batches were classified incorrectly (96.6% classified correctly). This re-substitution error rate can be interpreted as an optimistic (or biased) estimate of the true error rate, since we classified the same data that was used to estimate the logistic model. To obtain a less biased estimate, a jackknife procedure was used whereby the ith observation is classified by a model estimated from the remaining observations. The resulting rate of misclassification was 22%, where up to 25% is usually acceptable (Neter et al. 1990).

3.4. Optimizing Process Yield

Using the estimated logistic model in Equation (9), we now develop an optimization model for the pharmaceutical problem. From Equation (9), $E(Y_t) = \text{Prob}(Y_t = 1$, or high yield) and $[1 - E(Y_t)] = \text{Prob}(Y_t = 0$, or low yield). For T batches, we define a likelihood ratio, which is the ratio between the joint probability that T production batches are all low yield batches and the joint probability that T production batches are all high yield, as follows,

$$L = \frac{\prod_{t=1}^{T} [1 - E(Y_t)]}{\prod_{t=1}^{T} E(Y_t)}, \tag{10a}$$

$$= \prod_{t=1}^{T} \frac{[1 - E(Y_t)]}{E(Y_t)}, \tag{10b}$$

and, using Equation (9),

$$= \prod_{t=1}^{T} \exp\{h(X_{1t}, X_{2t}, \ldots, X_{Mt})\}. \tag{10c}$$

Note, that Equation (10b), and hence Equation (10c), represent a product of the odds ratios between low yield and high yield for each batch.

The objective is then to minimize the likelihood ratio (10a) by matching. With this objective, we seek to simultaneously minimize the joint likelihood of low yield batches, while maximizing the joint likelihood of

Table 2 Coefficient Estimates for the Logit Model*

Variable	Coefficient Estimate	Standard Error	Coefficient Variable	Standard Estimate	Standard Error
Intercept	1.2898	0.4492	X_2Z_{14}	−9.9855	2.3587
X_2Z_3	−3.5230	1.5279	$\boldsymbol{X_2X_8}$	**−5.1182**	**1.9796**
X_8Z_{12}	−1.6431	0.4538	Z_9Z_{13}	0.3498	0.1251
X_6Z_8	8.6850	2.9560	X_3	1.7091	0.5503
Z_2Z_3	−0.0142	0.00401	$\boldsymbol{X_1X_3}$	**−50.0991**	**13.2371**
Z_8Z_{14}	4.8083	1.5426	X_5	−12.6764	3.7253
Z_9	0.6872	0.2375	Z_2Z_8	0.00122	0.000503
X_4Z_5	0.0142	0.00449	Z_1Z_1	−0.6511	0.2496
Z_{13}	0.6297	0.1971	X_1Z_1	47.3706	14.0455
$Z_{12}Z_{14}$	−1.6426	0.4055	X_1Z_2	−0.1387	0.0472
$Z_{12}Z_{13}$	0.0818	0.0271	Z_2	−0.00477	0.00110
X_1Z_5	0.5759	0.1733	$\boldsymbol{X_1X_6}$	**−299.2**	**92.7867**
X_2Z_{11}	0.0671	0.0185	Z_{11}	0.0664	0.0188
X_7Z_{11}	−0.0772	0.0214	X_3Z_6	3.0106	1.1987
X_2Z_{10}	−0.8378	0.2046	$Z_{11}Z_{14}$	−0.4503	0.1198
X_8Z_5	0.0517	0.0196	X_8	−5.4234	2.4588
Z_{14}	6.0564	2.0989	X_1Z_3	112.8	57.2340
$Z_{10}Z_{11}$	−0.0333	0.00840	Z_8Z_{11}	−0.0280	0.0104
X_7Z_{10}	0.5634	0.1815			
$Z_{10}Z_{12}$	−0.1252	0.0347			
Z_3Z_{13}	−2.8772	1.1109			
$Z_{11}Z_{13}$	−0.0265	0.00873			
Z_3Z_{10}	−1.3048	0.6438			
Z_2Z_{11}	0.000117	0.000042			
X_5Z_1	19.5878	4.7258			
X_5X_5	−53.6386	15.6317			
X_7Z_1	−1.8430	0.5011			
X_7Z_4	2.4121	0.6310			
X_8Z_{13}	2.5103	0.9326			
$\boldsymbol{X_3X_5}$	**9.2587**	**4.1566**			

*Variables in bold are significant variables.

high yield batches. This can be accomplished by minimizing the natural log of Equation (10c) as follows,

$$Ln(L) = \sum_{t=1}^{T} h(X_{1t}, X_{2t}, \ldots, X_{Mt}), \tag{11}$$

where the expression in Equation (11) is the sum of the logs of the odds ratios for each batch. From the discussions in Example 2 and the development of Model (7) the only terms in the expected value of Equation (11) affected by matching are the interactive effects between variable means or $a_{ij}\mu_{ik}\mu_{jl}$. Hence the problem can be restated as follows,

$$\text{Min } E\left[\sum_{t=1}^{T} Q(Y_t(\{\delta_{ikjl}\}))\right] = \sum_{i=1}^{M}\sum_{j>i}^{M} a_{ij} \sum_{k=1}^{T}\sum_{l=1}^{T} \mu_{ik}\mu_{jl}\delta_{ikjl} \tag{12}$$

subject to Equations (5b) through (5e).

Problem (12) consists of T quadratic assignment problems. That is, a feasible solution to Equation (12) results in a quadratic assignment of raw materials for each of T batches. Further, the objective function in Equation (12) identifies the cost matrix as a *product matrix,* since the cost coefficient of δ_{ikjl} is the product of $\mu_{ik}\mu_{jl}$ (Plante et al. 1987). Using this property of the cost matrix, we now develop a procedure that, under special circumstances, obtains an optimal solution to Equation (12). For fixed i and j, any $T \times T$ matrix $\Delta_{ij} = (\delta_{ikjl})_{k,l=1,2,\ldots,T}$ which satisfies Equations (5b) through (5e) is a permutation matrix. Let $\phi_{ij} = (\phi_{ij}(1), \phi_{ij}(2), \ldots, \phi_{ij}(T))$ be a T vector with $\phi_{ij}(k) = l$, when $\delta_{ikjl} = 1$. Since each row and column of Δ_{ij} has exactly one "1", ϕ_{ij} represents a permutation of $1, 2, \ldots, T$. Note that, for every k, $\delta_{ikjl} = 1$ for exactly one index l, the mapping $\phi_{ij}(k) = l$ is well-defined. Given any matrix Δ_{ij} which satisfies Equations (5b) through (5e) and letting ϕ_{ij} be the corresponding permutation representation of Δ_{ij}, we note that

$$\sum_{i=1}^{M} \sum_{j>i}^{M} a_{ij} \sum_{k=1}^{T} \sum_{l=1}^{T} \mu_{ik}\mu_{jl} = \sum_{i=1}^{M} \sum_{j>i}^{M} a_{ij} \sum_{k=1}^{T} \mu_{ik}\mu_{j\phi_{ij}(k)}. \tag{13}$$

Without loss of generality, we assume that the mean values of variable i are in nonincreasing order $\mu_{i1} \geq \mu_{i2} \geq \ldots \geq \mu_{iT}$. Let ϕ^1_{ij} represent the permutation, such that the batch mean values of variable j are in the *same order* as those of variable i. In this case $\mu_{j,\phi^1_{ij}(1)} \geq \mu_{j,\phi^1_{ij}(2)} \geq \ldots \geq \mu_{j,\phi^1_{ij}(T)}$. Further, let ϕ^2_{ij} represent the permutation, such that the mean values of variable j are arranged in the *opposite order* as those of variable i. In this case $\mu_{j,\phi^2_{ij}(1)} \leq \mu_{j,\phi^2_{ij}(2)} \leq \ldots \leq \mu_{j,\phi^2_{ij}(T)}$. Thus, for this case, ϕ^1_{ij} represents a permutation of the mean values for the jth variable in nonincreasing order and ϕ^2_{ij} represents a permutation of the mean values for the jth variable in nondecreasing order.

To further facilitate the discussion of interaction effects on yield and the generation of yield maximizing solutions, let us define a graph, G, consisting of (1) nodes for each raw material characteristic (# nodes $= M$), and (2) edges connecting the nodes that have significant interaction effects (# edges $\leq \binom{M}{2}$). Further, a $(+1)$ value is assigned to an edge, say the edge connecting parameters i and j, where the sign of a_{ij} is such that a ϕ^1_{ij} permutation is suggested (mean values of the jth variable are sorted in the same direction as the mean values of the ith variable), otherwise a (-1) value is assigned. Define an *odd loop* as a cycle in G, where the product of the edge values on the cycle is (-1). When an odd loop does not exist, then there are no contradictory matches in the system and an optimal solution to Problem (12) can be obtained by the optimal matching of each individual pair represented by an interaction effect. (Note: In general finding an odd loop or determining that there is no odd loop cannot be done in polynomial time. However, for this case, and most other cases that we have encountered, where we are only interested in statistically significant interactions for matching, the assessment of whether odd loops exist is easily accomplished by hand. This is especially true since the number of nodes and edges required to determine the existence of odd loops is drastically reduced from $\binom{M}{2}2T$ to M nodes and from $\binom{M}{2}T^2$ to $\leq \binom{M}{2}$ edges, as previously discussed.)

If an odd loop does not exist, then the *minimization* problem in Equation (12) can be solved optimally using the following theorem.

THEOREM. *For each ij pair, where $a_{ij} > 0$, Equation (12) is minimized by the following optimal permutation,*

$$\phi^*_{ij} = \phi^2_{ij}, \tag{14}$$

and, for each ij pair, where $a_{ij} < 0$, Equation (12) is minimized by the following optimal permutation,

$$\phi^*_{ij} = \phi^1_{ij}. \tag{15}$$

PROOF. Given that the ijth component of Equation (13) does not belong to an odd loop in G, then we can optimize this component independently of the remaining components of Equation (13). Indeed, if there are no odd loops in G, then Equation (13) is optimized by independently determining the optimum solution for each component.

To show that Equation (14) is optimal for minimizing Equation (12) when $a_{ij} > 0$, it suffices to show that Equation (13) is nondecreasing for any permutation $\phi_{ij} \neq \phi^2_{ij}$.

Define a permutation ϕ_{ij} as follows $\forall a < b$,

$$\phi_{ij}(a) = \phi^2_{ij}(b),$$

$$\phi_{ij}(b) = \phi^2_{ij}(a),$$

and,

$$\phi_{ij}(k) = \phi^2_{ij}(k), \forall k \neq a, k \neq b.$$

The change in Equation (13) resulting from this permutation is then,

$$\delta = (\mu_{ia} - \mu_{ib}) \times (\mu_{j\phi^2_{ij}(b)} - \mu_{j\phi^2_{ij}(a)}).$$

By definition we have $\mu_{ia} \geq \mu_{ib}$, $\forall a < b$, and $\mu_{j\phi_{ij}^2(b)} \geq \mu_{j\phi_{ij}^2(a)}$, $\forall a < b$, which implies that $\delta \geq 0$, $\forall a < b$. Thus, any permutation $\phi_{ij} \neq \phi_{ij}^2$ results in a nondecreasing Equation (13), for $a_{ij} > 0$. Thus ϕ_{ij}^2 is the optimal permutation when $a_{ij} > 0$. A similar proof can be obtained for Equation (15), when $a_{ij} < 0$. □

If one or more odd loops exist then Duffy (1994) suggests the use of the simulated annealing algorithm as a heuristic for obtaining good solutions to Problem (12). This algorithm has worked well in many combinatorial optimization problems ranging from the Traveling Salesman problem to DNA mapping (Collins 1988). Further, the above solutions in Equations (14) and (15) can be used to provide an initial feasible solution for this algorithm.

As shown in Figure 2(a), no odd loops exist among the raw material quality characteristics, as there does not exist a cycle in G. Thus, the optimal matching of raw materials to minimize the likelihood ratio (or to minimize Equation (12)) can be obtained from Equation (14) or Equation (15).

For example, since $a_{28} = -5.1182 < 0$ for X_2 and X_8 (Table 2), then to reach the minimum in Equation (12), the effect of this coefficient is *maximized* by the following permutation to match batches of materials 2 and 8:

$$\phi_{28}^* = \phi_{28}^1.$$

Similarly, since $a_{13} = -50.0991 < 0$ for X_1 and X_3, then

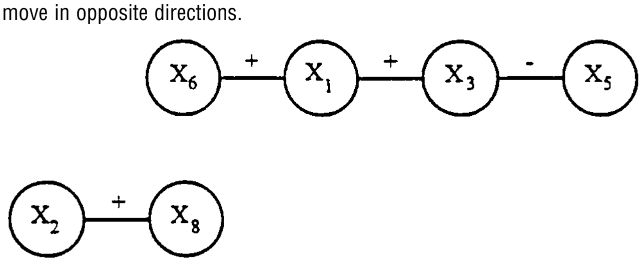
$$\phi_{13}^* = \phi_{13}^1.$$

Since $a_{35} = 9.2587 > 0$ for X_3 and X_5, then to minimize Equation (12), the effect of this coefficient is *minimized* by the following permutation,

$$\phi_{35}^* = \phi_{35}^2.$$

Figure 2a Graph of Raw Material Interactions

A (+) sign indicates that the connected variables should move in the same direction to improve process yield. A (−) sign indicates that they should move in opposite directions.

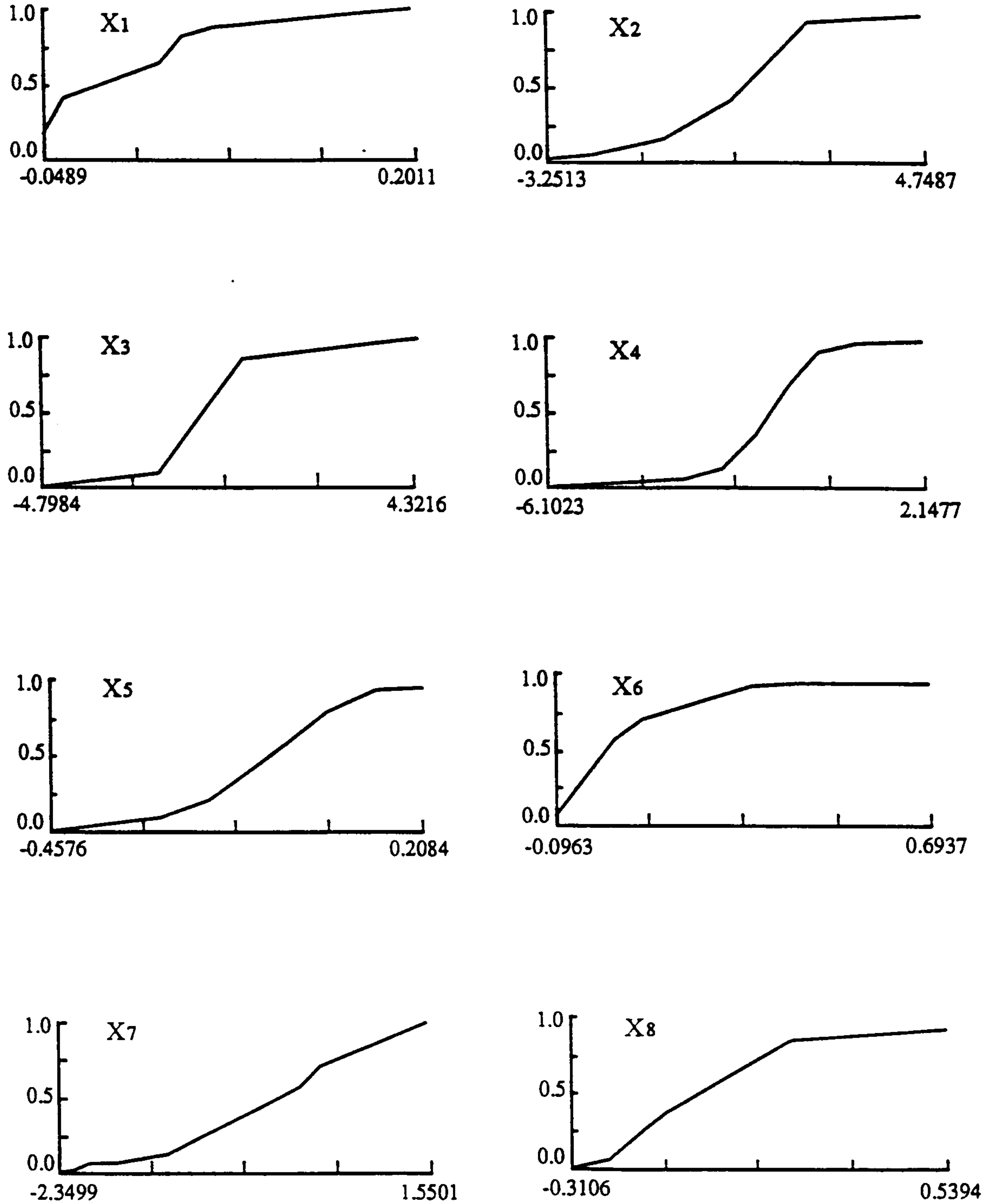


3.5. Simulation Analysis of Impact of Matching on Process Yield

The matching strategy from the previous sections could be immediately implemented on the raw materials. However, there remained practical barriers (inventory space related) to implementation that, in collaboration with the company, were not considered insurmountable. Prior to such an undertaking, the company wished to validate, via simulation, the matching approach and also consider potential long-term strategies for yield improvement. For these assessments, we considered the following three cases for a simulation study. Each of the simulation cases was constructed to illustrate the potential impact of matching.

In Case 1, which was the current situation, we simulated each of eight raw material process inputs using the current sample cumulative distributions shown in Figure 3. This case will illustrate the impact of matching without any change in supplier or manufacturer process capabilities. In Case 2, we considered that knowledge concerning supplier raw material processes had increased (improved) to the extent that they have the capability to produce output that is either one standard deviation below $(\mu_i - \sigma_i)$ or one standard deviation above $(\mu_i + \sigma_i)$ the mean for each raw material characteristic. Further, given this improved capability, the output of the raw material processes [either $(\mu_i - \sigma_i)$ or $(\mu_i + \sigma_i)$] was *prespecified* by the manufacturer. This case will illustrate the impact of matching given that suppliers have improved their process capabilities without concurrent improvement in manufacturer capabilities. Case 3 extends Case 2 by also considering the effect of controlling some of the manufacturer's parameters that are currently considered uncontrollable, or possibly too costly to control. The selected process process control parameters were Z_4, Z_5, Z_6, Z_7, and Z_8. To illustrate the potential impact of controlling these process parameters, adjustments for these parameters were selected from a set of two simulated settings that consisted of either one standard deviation below the mean $(\mu_m - \sigma_m)$ or one standard deviation above the mean $(\mu_m + \sigma_m)$. Information obtained from this case will be used to provide guidance

Figure 3 Standardized Sample Cumulative Probability Function for Each Raw Material



for the development of long-term strategies for improving process yield. (The results of this case should provide information about the relative impact of considering process control adjustments as well as provide information concerning the appropriate adjustment levels for each process control parameter. For purposes of this case, we have assumed that a set of T possible process control parameter settings is given. In practice, however, such adjustments are not given a priori, but should be simultaneously determined when solving the raw material matching problem. An extension to the mathematical formulation for the matching problem, having appropriate operational range constraints for the adjustment of each process control parameter, and the development of corresponding solution strategies are the subjects of our ongoing research efforts in this area.)

For each of the above three cases, four values for the number of production batches were used ($T = 4, 8, 12$, and 16). In Case 1 for each of the 8 raw materials considered, T values were simulated from their respective empirical cumulative distribution in Figure 3 for the T batches of each raw material. Recall in Case 1, the process parameters were considered uncontrollable; but each could assume one of the two possible values (high and low). Thus, a production batch with a set of values on $X_1 - X_8$ for the eight raw material batches would actually be processed under one of 32 possible process control parameter value combinations, and hence gave 32 possible $E(Y)$s via Equation (9). (We used a fractional factorial design over the 14 process parameters, such that there were $2^{14-9} = 2^5 = 32$ design points, where Z_4, Z_5, Z_6, Z_7, and Z_8 were the base variables used for constructing the design.) Assuming that the high and low process parameter values were equally likely, we computed $E(Y)$ for each of these 32 combinations, and used the average of $E(Y)$ over these 32 combinations as the likelihood of a high process yield for this set of raw material inputs. In Cases 2 and 3, for each raw material characteristic and process parameter, exactly one half of the T values were prespecified at $(\mu_i + \sigma_i)$, with the remaining at $(\mu_i - \sigma_i)$. To assess the impact of matching on yield, the likelihood of high yield was determined via Equation (9) for the proposed matching method, as well as random selection. We used 100 replications of the above simulation and computed the overall average probability of high process yield. For comparative assessment, this average probability was recorded for both procedures as well as the percentage improvement of the proposed matching method over random selection (Table 3).

Since there were no odd loops (Figure 2a), raw material matching from Equation (14) or Equation (15) gave the solution which minimized Equation (12) for Cases 1 and 2.

In addition, as shown in Figure 2b, there were no odd loops in the graph representing the interaction effects among the raw materials and selected process control parameters (Case 3). Thus matching in Equations (14) or (15) was again optimal for minimizing Equation (12).

The results shown in Table 3 are revealing in terms of the potential benefits derived from the informed use of matching. For example, even with increasingly more control of supplier (raw material) processes (Cases 2 and 3) and manufacturer's processes (Case 3), the continued use of random selection of raw material batches (due to prior beliefs that raw material levels did not

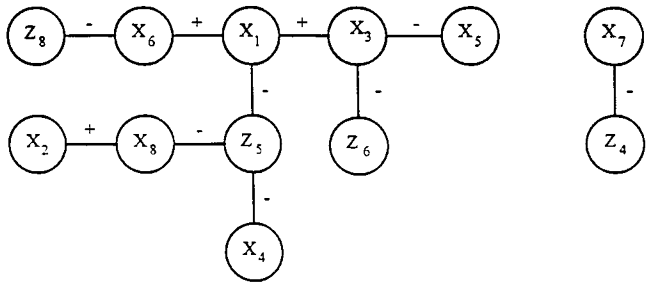
Table 3 Process Simulation Results of Pharmaceutical Example*

T	Probability of High Yield: Random (Uninformed Matching)	Probability of High Yield: Optimal (Informed Matching)	% Improvement
Case 1: Current Supplier Technology			
4	0.5215	0.5630	7.968
8	0.5265	0.5806	10.261
12	0.5271	0.5838	10.745
16	0.5260	0.5873	11.643
Case 2: Improved Supplier Technology			
4	0.5304	0.6619	24.775
8	0.5254	0.6619	25.975
12	0.5235	0.6619	26.428
16	0.5250	0.6619	26.064
Case 3: Improved Supplier and Manufacturer Technology			
4	0.4958	0.99996	101.681
8	0.5006	0.99996	99.761
12	0.5254	0.99996	90.310
16	0.5028	0.99996	98.858

*A 1% improvement in the average yield of this product was estimated to provide a one million dollar return increase for the company.

Figure 2b Graph of Raw Material and Some Process Variable Interactions

A (+) sign indicates that the connected variables should move in the same direction to improve process yield. A (−) sign indicates that they should move in opposite directions.



affect product performance) completely masks the potential gains that are possible via raw material matching. Indeed, the lowest probabilities occurred under the case of greatest process capability improvement (Case 3), where random selection achieved a likelihood of only 0.4958 for $T = 4$. The results in Table 3 clearly demonstrate the extent of process improvement that can be derived from matching. The percent improvement of matching over random selection ranges from a low of about 8% for the immediate case of no process improvements to more than 100% for the case of supplier and manufacturer process improvement. In the latter case the likelihood of achieving a high yield batch is essentially 1.0. Recall that a 1% improvement in the average yield of this product was expected to provide a one million dollar return for the company.

Furthermore, as the number of batches (T) increases, the likelihood of achieving a high yield increases when the raw material input is random (Case 1). However, when one specifies the incoming raw material batches (Case 2), and further matches, via adjustment, selected process control parameters (Case 3), then the likelihood of a high yield batch is not affected by the number of batches, T.

3.6. Long-Term Strategies for Improving Yield

Given the clear and substantial benefits that the simulations showed, company representatives decided to (1) immediately implement, where feasible, the matching, via adjustment, of process controls to raw material batches, and (2) pursue an investigation into the feasibility of sorting raw material batches and adjusting process variables accordingly. The prevailing feeling was that it would not be practical to sort the batches on site. Rather, using the simulation results as guidance, improving supplier processes was considered a more viable alternative. To this end, strategies for long-term improvement were developed.

Although, for purposes of empirical assessment of matching, two raw material and process outputs ($\mu_i - \sigma_i$ and $\mu_i + \sigma_i$) were considered possible in Cases 2 and 3, it is more likely that only one output without variation is specified for a given raw material process (i.e., $\mu_i + k_i\sigma_i$, where the sign on k_i can be determined by informed matching). Reducing and eventually eliminating variability should be a main goal for long-term improvement. In such precise cases, matching aids the manufacturer in pursuing a yield improvement strategy that removes the need for sorting raw material batches. Figure 2b gives only the optimal pairwise matchings of individual pairs; but there are four possible strategies with $k_i = \pm 1$ for all i, that satisfy the optimal pairwise matchings depicted in Figure 2b. Since there are $c = 2$ unconnected graphs in Figure 2b, there are $2^c = 2^2 = 4$ possible supplier/process strategies as follows.

Strategy 1. *Raw Material (X):* ($\mu_6 + \sigma_6$, $\mu_1 + \sigma_1$, $\mu_3 + \sigma_3$, $\mu_5 - \sigma_5$, $\mu_8 + \sigma_8$, $\mu_2 + \sigma_2$, $\mu_7 + \sigma_7$, $\mu_4 + \sigma_4$). *Process Controls (Z):* ($\mu_8 - \sigma_8$, $\mu_6 - \sigma_6$, $\mu_5 - \sigma_5$, $\mu_4 - \sigma_4$).

Strategy 2. *Raw Material (X):* ($\mu_6 + \sigma_6$, $\mu_1 + \sigma_1$, $\mu_3 + \sigma_3$, $\mu_5 - \sigma_5$, $\mu_8 + \sigma_8$, $\mu_2 + \sigma_2$, $\mu_7 - \sigma_7$, $\mu_4 + \sigma_4$). *Process Controls (Z):* ($\mu_8 - \sigma_8$, $\mu_6 - \sigma_6$, $\mu_5 - \sigma_5$, $\mu_4 + \sigma_4$).

Strategy 3. *Raw Material (X):* ($\mu_6 - \sigma_6$, $\mu_1 - \sigma_1$, $\mu_3 - \sigma_3$, $\mu_5 + \sigma_5$, $\mu_8 - \sigma_8$, $\mu_2 - \sigma_2$, $\mu_7 + \sigma_7$, $\mu_4 - \sigma_4$). *Process Controls (Z):* ($\mu_8 + \sigma_8$, $\mu_6 + \sigma_6$, $\mu_5 + \sigma_5$, $\mu_4 - \sigma_4$).

Strategy 4. *Raw Material (X):* ($\mu_6 - \sigma_6$, $\mu_1 - \sigma_1$, $\mu_3 - \sigma_3$, $\mu_5 + \sigma_5$, $\mu_8 - \sigma_8$, $\mu_2 - \sigma_2$, $\mu_7 - \sigma_7$, $\mu_4 - \sigma_4$). *Process Controls (Z):* ($\mu_8 + \sigma_8$, $\mu_6 + \sigma_6$, $\mu_5 + \sigma_5$, $\mu_4 + \sigma_4$).

Evaluating the likelihood of a high process yield for each of these supplier/process strategies will determine the appropriate long-term strategy for integrated supplier and manufacturer process improvement. The evaluation of each of these strategies is necessary, since the main effects, as well as interaction effects on expected yield, need to be considered in assessing a *strategy for improving yield*. Using Equation (6), Equation (9), and the results in Tables 1 and 2, the expected yields are 0.99942, 0.99543, 0.99997, and 0.99999, respectively for strategies 1, 2, 3, and 4. Thus, strategy 4 would be the preferred strategy. Further, depending on process capability and accumulated process knowledge, k_i can be adjusted to improve yield further. The direction for adjustment is determined by the sign $(\pm)$ of σ_i in the selected strategy. For example, if $k_i = 2$ for all raw material and process control parameters, the likelihood of high yield is 1.00000 for all strategies (1, 2, 3, and 4).

Clearly, the company representatives had come a long way from deciding that there was no impact of raw material inputs on process yield to an understanding and awareness of the potential benefits of improving raw material processes. However, as in many such projects of this nature, the company representatives that we worked with were all, at different times, transferred. Also, at this time, health care reform was high on the national agenda and the pharmaceutical industry was facing the very real possibility of price controls. Further, a new CEO had just been hired. Consequently, implementation of the results was put on hold.

Nonetheless, the message of this research was clear to the pharmaceutical company. Learning about, understanding, and taking advantage, via matching, of the effect of *interactions* between raw materials and process controls on process performance can provide significant advantages for short as well as for long term process improvement and hence product yields.

4. Discussion

A combinatorial optimization model was proposed for matching raw material batches over a specified inventory horizon, such that, in the presence of significant interaction effects, expected quality of a product is maximized. Motivating this model is the observation that raw material variation is often the most significant factor affecting product quality. Under certain conditions, the optimal solution can be obtained in a straightforward manner using the results of Plante et al. (1987). When such conditions are not met, then good solutions can be obtained via heuristic algorithms, such as the simulated annealing algorithm (Collins 1988). The formulation, and solution of the proposed models was demonstrated via a case study of a major pharmaceutical manufacturer. The results of these examples indicate that raw material matching can significantly improve the product quality of a manufacturer's process.

The empirical results also suggest a number of implications with regard to the planning and assessment of quality improvement strategies. In particular, when interactions exist among influential factors (both supplier output and manufacturer process controls), there is a substantial impact on quality improvement (increased process yields) obtained via matching. As the empirical results suggest, informed matching will always improve quality compared to the typical use of random selection of supplier batches. Furthermore, improving supplier capability and improving quality via informed matching with manufacturer process control parameters can be an exceptionally beneficial combination. Indeed, if matching possibilities are not known nor used, then knowledge and use of improved supplier technology may prove virtually worthless under the typical use of random selection, completely masking the benefits of such improvement efforts and frustrating the consideration of further improvement efforts.

When interactions among factors exist, matching further provides a useful planning tool for long-term strategic quality improvement. In particular, the manufacturer can work with suppliers in a more informed manner to develop supplier/manufacturer integrated strategies for quality improvement. In this regard the recognition or identification of yield maximizing nominal or target values for the means of supplier processes are readily available via matching. Suppliers can then concentrate on obtaining the necessary technology to achieve these specified targets and to reduce the process variance about these targets.[1]

[1]The authors gratefully acknowledge the assistance provided by the

References

Askin, R. A., J. B. Goldberg. 1988. Economic optimization in product design. *Engrg. Optimization* **14.**

Collins, N. 1988. Simulated annealing—An annotated bibliography. *Amer. J. Math. Management Sci.* **3.**

Duffy, J. 1994. Essays on product design, and manufacturing decisions. Unpublished Ph.D. dissertation, Purdue University.

Gabor, A. 1989. The front lines of quality. *US News and World Report* November 27, 57–59.

Kelly, K. 1993. Cut costs or else. *Business Week* March 22, 28–29.

——. 1994. Motorola: Training for the millennium. *Business Week* March 28, 158–163.

Editor-in-Chief, Senior Editor, and especially the two referees, who provided several valuable insights. Their comments and suggestions have resulted in a significantly improved manuscript. This research was supported, in part, by grants from the Center for the Management of Manufacturing Enterprises (CMME), Purdue University.

Marsh, C., T. Tucker. 1990. Application of SPC techniques to batch units. *ISA Trans.* **3** 41–46.

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

Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization.* John Wiley and Sons, New York.

Neter, J., M. H. Kutner, C. J. Nachtsheim, W. Wasserman. 1990. *Applied Linear Statistical Models*, 4th ed. Irwin, Chicago, IL.

Papadimitriou, C., K. Steiglitz. 1982. *Combinatorial Optimization.* Prentice-Hall, Englewood Cliffs, NJ.

Pignatiello, J. J. 1993. Strategies for robust multiresponse quality engineering. *IIE Trans.* **25** 3.

Plante, R. D., T. J. Lowe, R. Chandrasekaran. 1987. The product matrix traveling salesman problem: An application and solution heuristic. *Oper. Res.* **35** 5 772–783.

Taguchi, G. 1987. *System of Experimental Design.* Kraus International Publications, White Plains, NY.

Vining, G. G., R. H. Myers. 1990. Combining Taguchi and response surface philosophies: A dual response approach. *J. Quality Techn.* **22** 1.

Accepted by Timothy Lowe; received September 27, 1996. This paper has been with the authors 7 months for 4 revisions.