A Mahalanobis Taguchi System-based Approach for Correlated Multiple Response Process Monitoring

Sagar Sikder
Statistical Quality Control & Operation Research Unit
Indian Statistical Institute
Mumbai, 400020, India

Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology
Bombay-400076, India

Subhash Chandra Panja
Jadavpur University
Kolkata-700032, India

Sasadhar Bera
Indian Institute of Management (IIM)
Ranchi- 834008, India

Abstract

Developing a single control chart for monitoring correlated multiple quality characteristics (or responses) is always a challenging research endeavor. In this context, various classical approaches, such as χ^2 and Hotelling T^2 control chart are recommended by researchers. However, interpretation, understanding of out-of-control state, and developing a control strategy for such charts are quite difficult. Difficulty is significant due to lack of understanding (knowledge) at the operational (worker) level. Monitoring each response individually and overall concluding on in-control state may be erroneous in case there is evidence of correlation between responses. In this paper, an easy-to-interpret integrated approach, based on Mahalanobis Taguchi system (MTS), binary logistic regression and neural network, is proposed. The approach adopts MTS to define a normal group and the Mahalanobis space, considering Hotelling T^2 control limits. Subsequently using unknown observations and nonlinear regression approach, a systematic approach to define cutoff value is proposed to classify normal and abnormal group. Logistic regression model is observed to provide more conservative cutoff as compared to a neural network. The overall integrated approach is verified using a real life industry case. The results clearly indicate the suitability of the approach as compared to classical approach at operations level.

Keywords
Multiple quality characteristic, Hotelling T^2, Mahalanobis Taguchi System (MTS), binary logistic regression, neural network.

1. Introduction
Woodall and Montgomery (1999) stated that multivariate statistical process control (MSPC) is one of the most rapidly developing areas of statistical process control. Nowadays, in industry, there are many situations in which the simultaneous monitoring or control of two or more related quality-process characteristics is important and necessary. Montgomery and Mastrangelo (1991) stated that monitoring of these quality characteristics independently can be very misleading. Thus, simultaneous process monitoring tool for several related characteristics is necessary. Monitoring and controlling several quality characteristics (responses) in a single chart is known as multivariate statistical process control. Different types of multivariate quality control charts are the primary tool for multivariate

statistical process control (Bersimis et al. 2006). In this regard, Hotelling T^2 is one of the recommended control chart for MSPC. In a discussion paper Jackson (1985) describes the application of Hotelling T^2. In a review paper, Lowry and Montgomery (1995) showed that for small sample size (<100 samples), Hotelling T^2 provides more accurate result compared to another multivariate control chart, known as χ^2 chart. . Hotelling T^2chart can also detect out of control signal quickly if the magnitude of process mean shift is greater than 3. Tracy et al. (1992) advocated the use of beta distribution for defining exact control limits in stage-1 of T^2 chart. It is to be mentioned that one of the basic assumption for Hotelling T^2chart is that the p-dimensional process characteristic $X = (X_1, X_2, \ldots, X_p)'$ follows multivariate normally distribution. In many practical situations, this assumption may be violated. Taguchi and Rajesh (2000) proposed a distribution free diagnostic and forecasting approach known as Mahalanobis-Taguchi System (MTS) for multivariate data analysis. Woodall (2003) provide a critical review on MTS and its limitations. However, in one of the book, Taguchi and Rajesh (2002) mentioned that MTS can be successfully applied as an alternative procedure for control charting for multivariate process control. MTS is a distribution free approach to identify out-of-control signal in multivariate monitoring scenario. MTS can also detect independent variables responsible for out-of-control state. MTS works on first identifying a group of items which are considered as healthy group. Healthy or normal group is used to create the Mahalanobis Space (MS) for reference. Any future observation which is outside the MS or having large Mahalanobis Distance (MD) is considered as abnormal or an out-of-control observation. However, identification of the normal group requires domain expertise. Also the boundary of the MS is not clearly specified or defined. In this context, recently, Wang et al. (2013) discusses on MS construct and its effectiveness on MTS strategy.

Instigated by the advantages and limitation of Hotelling T^2 and MTS, an integrated and simplified nonlinear approach, which can be implemented in shop floor, is proposed in this paper. However, in the proposed approach, Hotelling T^2 chart is used as reference to identify a normal group that can be used to develop the Mahalanobis space (MS). Subsequently, binary logistic regression is used to explore the relationship between MD and binary response (control or out-of-control state) to identify a boundary value for the MS. A nonlinear neural model is also developed, verified, which is equally efficient to identify the state of control, and easy to use for real time control.

A brief detail on the techniques used for the proposed integrated monitoring approach is provided in the following section.

2. Multivariate Techniques

Let us suppose that X is a $p \times 1$ random vector, where $X_1, X_2, \ldots, X$, each representing p different quality characteristics to be monitored over time.

2.1. Hotelling T^2 control chart

If vector X follows normal distribution, it can be expressed as $X \approx N_p(\mu_0, \Sigma_0)$. The test statistic for the i-th individual observation is

$$D_i^2 = (X_i - \mu_0)' \Sigma_0^{-1} (X_i - \mu_0) \tag{1}$$

where, X_i : *i*-th observation, *i*=1,2,…,*m*, μ_0 : Mean Vector, $\mu_0 = (\mu_1, \ldots \mu_p)'$ and Σ_0 : known variance-covariance matrix. Here it is assumed that the observations X_i are independent. Thus, D_i^2 statistic follows a χ^2 distribution with p degrees of freedom (Montgomery 2001).Thus, with known mean vector, μ_0 , and variance-covariance matrix Σ_0 the χ^2 upper control limit is defined as $L_u = \chi^2_{p,1-\alpha}$. If μ_0 is replaced by $\overline{X_0}$ (i.e. sample mean vector) and Σ_0 is replaced by S_0 (i.e. Sample variance covariance matrix) and X_i is the *i*-th individual observation, then $T_i^2 = (X_i - \bar{X}_0)' S_0^{-1} (X_i - \bar{X}_0)$ statistic is known as the Hotelling T^2 and follows a β -distribution with *p*/2 and (*m*- *p* -1) degrees of freedom, Thus a multivariate Hotelling T^2 with unknown parameters has the following upper control limit (Tracy et al. 1992)

$$L_u = \frac{(m-1)^2}{m} \beta_{1-\alpha, p/2, (m-p-1)/2} \tag{2}$$

where m denotes the total number of observations, and α is the level of significance.

2.2 Mahalanobis Taguchi System (MTS)

The MTS (Taguchi and Rajesh 2002) is proposed as a diagnosis and forecasting approach for various multivariate data. In this approach, data must be available on 'healthy' or 'normal' group of items, so-called 'reference group'. There is also need for number of 'abnormal' items, classified based on severity levels of abnormalities. One of the main objectives of MTS is to introduce a scale-based on all input characteristics to measure the degree of abnormality. In MTS, Mahalanobis distance (MD) is used as a measurement scale. MTS can be divided into into four stages. In stage 1, the variables that define the "healthiness" of an item are identified. Data are collected on the healthy or normal group. The variables are standardized and the MD's are calculated for the normal group. These values define the "Mahalanobis Space" used as a frame of reference for the MTS measurement scale.

We refer to the variables collected on each item to determine its "healthiness" as $X_i, i=1,2,...,p$. We denote X_{ij} as observation of the i-th variable on the j-th item, where i=1,2,…,p, j=1,2,…,m. Thus the $p\times 1$ data vectors for normal group are denoted by $X_j, j=1,2,...,m$; m being the total number of observations.

Each individual variable in each data vector is standardized by subtracting the mean of the variable and dividing by its standard deviation, with both statistics calculated using data on the variable in the normal group. Thus we have the standardized values (Z_{ij}) is given by

$$Z_{ij} = (X_{ij} - \overline{X_i})/S \ , \qquad i = 1,2,...,p, \quad j = 1,2,...,m \tag{3}$$

where

$$\overline{X_i} = \sum_{j=1}^{m} X_{ij} \Big/ m \tag{4}$$

and

$$S_i = \sqrt{\sum_{j=1}^{m} (X_{ij} - \overline{X_i})^2 \Big/ (m-1)} \ . \tag{5}$$

Next the values of the MDs, $MD_j, j=1,2,...,m$, are calculated for the normal items using

$$MD_j = (1/p) Z_j^{'} S^{-1} Z_j \ , \tag{6}$$

where $Z_j^{'} = [Z_{1j,} Z_{2j},...,Z_{pj}]$ and S is the sample correlation matrix expressed as

$$S = 1/(m-1)\sum_{j=1}^{m} Z_j Z_j^{'} . \tag{7}$$

Taguchi and Rajesh (2002) claim that MD's calculated using Eqn. (6) have an average value of unity or 1. For this reason they also refer to the Mahalanobis space (MS) as the 'unit' space.

In stage 2, abnormal items are selected. There is no uncertainty involved in the MTS regarding the status of each item used for determining the MTS measurement scale. The MD's of the abnormal item with data vectors denoted by $X_{j,} j = m+1, m+2,..., m+t$ are calculated after the variables are standardized using the normal-group means and standard deviations. Thus we have $MD_j, j = m+1, m+2,..., m+t$ with MD_j as defined in Eqn. (4) and Eqn. (5). According to MTS, the resulting MD scale is satisfactory if the MD_j values for the abnormal items are quite higher than those for the normal items.

In stage 3, orthogonal arrays (OA) and S/N ratios are used to identify the most useful set of variables. In final stage, a confirmation run is conducted based on useful set of variables. Stage 3 and Stage 4 is not used for this integrated approach. The following case study shows the step by step approach followed to develop a nonlinear, and easy-to-use monitoring tool. However, the reference point for developing the approach is Hotelling T^2 and MTS.

3.0 Verifying the Proposed Approach using Case

Mukherjee and Ray (2006) collected time oriented multivariate data for finish grinding. Total *198* observations for five critical response variables, viz. average surface finish (Y_1), honing angle(Y_2), average liner bore diameter (Y_3), maximum bore ovality (Y_4) and maximum bore taper (Y_5). To develop the proposed approach, we considered *43* points to define the in-control state based on Hotelling T^2 and also shown in Figure 1.

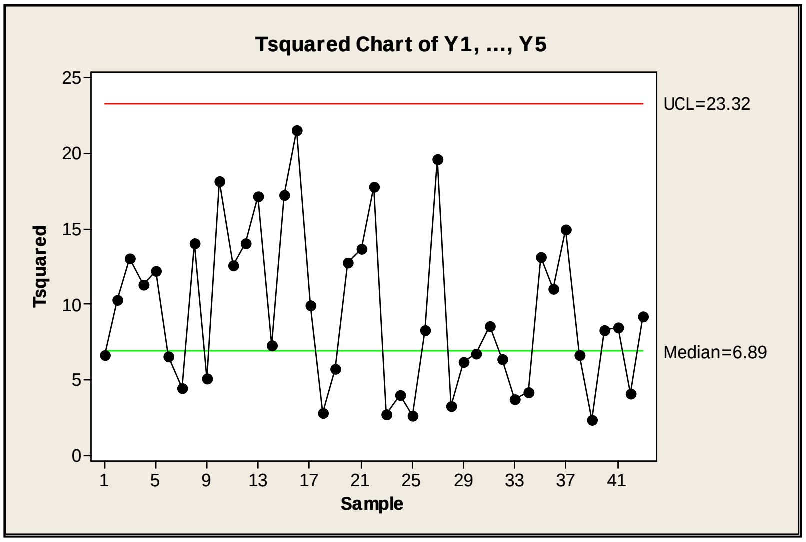

Figure 1: In-control observation in Hotelling T^2 control chart

The observations, as shown in figure 1 are used to construct the UCL for the Hotelling T^2 chart as well as to construct the Mahalanobis Space (MS). The mean vector of the *43* observations is given below. This vector is used to calculate the T^2 statistic in Hotelling T^2 chart.

$$\overline{Y} = \begin{bmatrix} 2.578 \\ 44.1744 \\ 97.0084 \\ 0.003 \\ 0.0035 \end{bmatrix} \tag{8}$$

The correlation coefficients of the responses are provided in Table 1. The matrix indicates significant correlation and a typical case of multivariate data.

Table 1: Correlation Matrix

Responses	y1	y2	y3	y4	y5
y1	1.000	-0.054	0.530	0.215	0.387
y2	-0.054	1.000	-0.583	-0.058	-0.073
y3	*0.530	*-0.583	1.000	0.104	0.048
y4	0.215	-0.058	0.104	1.000	0.488
y5	*0.387	-0.073	0.048	*0.488	1.000

*Significant correlation at *5* % level of significance

The correlation matrix is used to determine the inverse correlation matrix, which is used to calculate MD's for the MS. The Inverse correlation Matrix is shown in Table 2.

Table 2: Inverse Correlation Matrix

Responses	y1	y2	y3	y4	y5
y1	2.103	-0.869	-1.589	0.070	-0.835
y2	-0.869	1.881	1.544	-0.078	0.438
y3	-1.589	1.544	2.729	-0.188	0.688
y4	0.070	-0.078	-0.188	1.327	-0.671
y5	-0.835	0.438	0.688	-0.671	1.650

Subsequently, the average MD for the "normal group" is calculated as *0.97732*, which is very close to unity. This indicates that the construction of MS is adequate. To calculate $T^{2,}$ variance covariance matrix is needed. The variance-covariance matrix based on *43* observations is given in Table 3. The inverse of variance-covariance matrix given in Table 3 is provided in Table 4.

Table3: Variance -Covariance matrix for the in-control observations.

Responses	y1	y2	y3	y4	y5
y1	0.2173	-0.0317	0.0009	0.0001	0.0003
y2	-0.0318	1.5671	-0.0027	-0.0001	-0.0001
y3	0.0009	-0.003	0.0000	0.0000	0.0000
y4	0.0001	-0.0001	0.00000	0.0000	0.0000
y5	0.0003	-0.0001	0.0000	0.0000	0.0000

Table 4: Inverse of variance-covariance matrix

Responses	y1	y2	y3	y4	y5
y1	9.5693	-1.4554	-906.0506	96.241	-1190.707
y2	-1.4554	1.1902	329.359	-37.684	228.7540
y3	-906.0506	329.359	198054.433	-30346.045	121871.5233
y4	96.2415	-37.684	-30346.044	434104.57	-266455.131
y5	-1190.707	228.754	121871.523	-266455.13	754793.0581

The inverse of variance- covariance matrix is used to calculate the T^2 statistic for Hotelling T^2 chart. The upper control limit for the T^2 chart is given by

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha/2, p, m-p} \quad (9)$$

Where, m = number of observations, p = number of characteristics, F indicates that the F distribution is used. In our case, m=*43*, p=*5* , and α = *0.27*, and UCL calculated is *28.24*. The T^2 chart is shown in Figure 2.

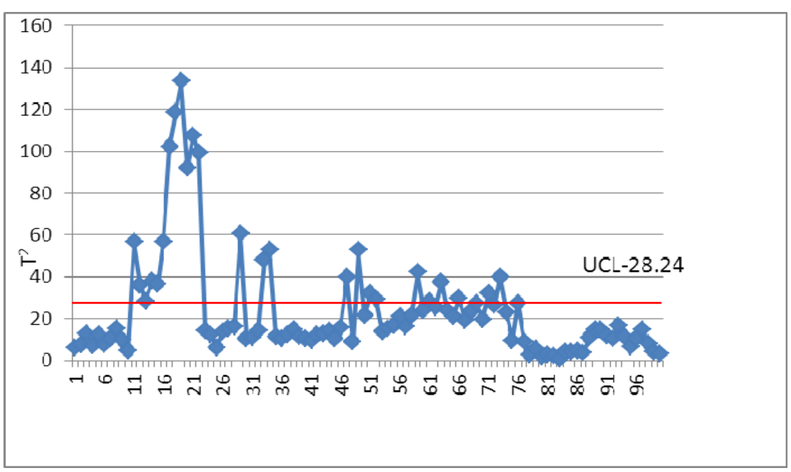

Figure 2: T^2 chart using 'm' observations

The above chart is used to detect the in-control and out-of-control signal for *98* new observations. The MD for each observation is also calculated for subsequent analysis. As, we want to use Mahalanobis distance as a measure to access control state, we need some threshold value, which will differentiate control and out-of-control state. Thus, binary logistic regression model is used to identify an appropriate threshold value. MD values of *98* observations are regressed with the binary states (defined as *1* for out of control and *0* for in control) as obtained from Hotelling T^2 chart. Table 5 shows the threshold MD as *5.713*, which differentiate in control and out of control state based on logistic regression model. Only the relevant parts of *100* observations are used in table 5.

Table 5: The MD values and corresponding probability of out-of-control state.

MD	Binary Response (based on Hotelling T^2)	Probability of being 1(out of control)
4.839	0	0
5.146	0	0
5.315	0	0
5.472	0	2.88E-07
5.498	0	5.69E-06
5.713	1	0.999
5.769	1	1
5.912	1	1
6.010	1	1
6.469	1	1
6.538	1	1

Any MD value greater that the limiting value will be considered as out-of-control point. Another nonlinear approach, using neural net (based on Gauss Newton algorithm), based on same 100 data set, is used to recalculate cutoff MD. Figure 3 shows network model, where Hi is the hidden node, MD is the input and T2 result is the binary response from Hotelling T^2 chart.

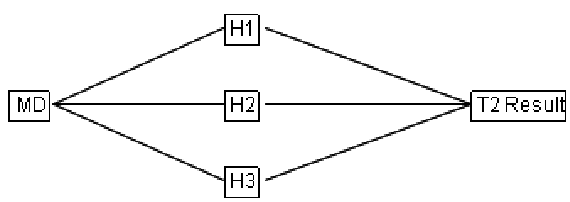

Figure 3: Neural network to map MD and Binary T2 output

The network parameters selected are, *3* hidden nodes, *0.01* over fit penalty, *16* numbers of tours (rerun and fit *16* times), *100* iteration, and convergence criteria is 0.00001. The R2 value for the fit is *0.9577* and SSE (sum of square error) is *4.18*. The cut-off MD based on maximizing a desirability function is *6.851*. The prediction profile is shown in Figure 4. Comparing the nonlinear models, neural network may result in more false alarm as its cutoff MD is higher than the logistic regression-based cutoff MD.

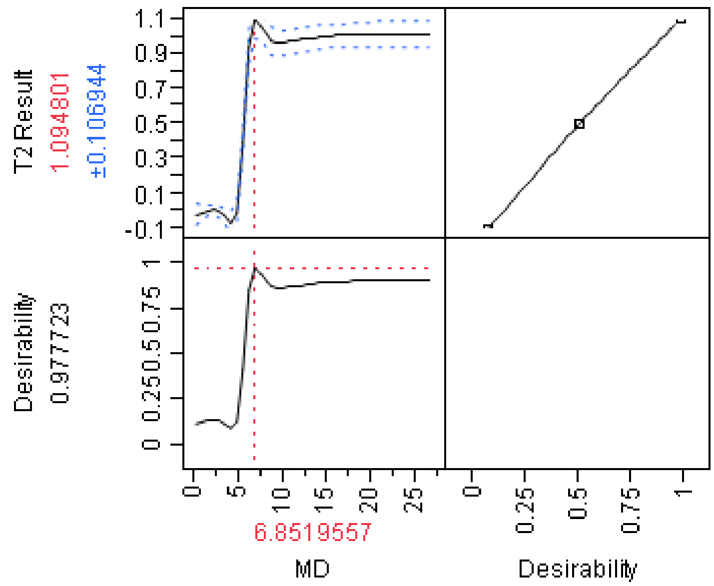

Figure 4: Neural Network based cutoff MD

Thus in this paper, a systematic approach to define a cutoff for MD based on MTS approach is illustrated. Nonlinear logistic regression seems more adequate as compared to neural net. However fine tuning of neural network may improve the solution. The overall integrated approach attempts to simplify the classical monitoring approach for multivariate responses. Extending the concept of MTS with defined approach for cutoff may lead to correctly identification the influential variables.

Conclusions

In this paper, an integrated approach to process monitoring of multiple response characteristics is proposed and verified using real life case example. The Mahalanobis Taguchi System (MTS) approach is first mapped with Hotelling T^2 approach to identify the normal or in-control group. Subsequently, a Mahalanobis space is created with the normal group. To define the cutoff, unknown observations and corresponding MDs are calculated. The cutoff for MD (to classify normal and abnormal observations) is based on nonlinear mapping of these MDs with binary response from Hotelling T^2 chart. Nonlinear regression model (logistic regression and artificial neural network) are preferred for selecting the cutoff. Logistic regression is observed to provide conservative cutoff value as compare to neural network. However, more data is to be collected to validate the defined cutoff. In addition, future scope exists for researchers to work on real-time identification of critical or influential independent variables, after validating the cutoff, for better process control.

References

Bersimis, S., Psarakis, S., and Panaretos, J., Multivariate Statistical Process Control Charts: An Overview. *Quality & Reliability Engineering International*, vol. 23, no. 5, pp. 517-543, 2006.

Jackson, J.E., Multivariate quality control. *Communication in Statistics-Theory and Methods*, vol. 14, pp. 2657-2688, 1985.

Lowry, C.A., and Montgomery, D.C., A review of multivariate control charts. *IIE Transactions*, vol. 27, pp. 800-810, 1995.

Montgomery, D.C., *Introduction to statistical quality control*, 4th Edition, John Wiley & Sons., New York, 2001.

Montgomery, D.C., Mastrangelo, C.M., Some statistical process control methods for autocorrelation. *Journal of Quality Technology*, vol. 23, pp. 173-193, 1991.

Mukherjee, I., and Ray, P.K., A review of optimization techniques in metal cutting processes, *Computer and Industrial Engineering*, vol. 50, pp. 15-34, 2006.

Taguchi, G., and Rajesh, J., New Trends in Multivariate Diagnosis, Sankhya, vol. 62, pp. 233-248, 2000.

Taguchi, G., and Rajesh, J., *The Mahalanobis-Taguchi Strategy A Pattern Technology System*, John Wiley & Sons, New York, 2002.

Tracy, N.D., Young, I.C., and Mason, R.L., Multivariate control charts for individual observations. *Journal of Quality Technology*, vol. 24, pp. 88-95, 1992.
Woodall, W.H., and Montgomery D.C. Research Issues and Ideas in statistical process control. *Journal of Quality Technology*, vol. 31, pp. 376-386, 1999.
Wang, N. , Saygin, C, and Sun, S., Impact of Mahalanobis space construction on effectiveness of Mahalanobis-Taguchi system, *Int. J. of Industrial and Systems Engineering*, vol.13, no.2, pp.233 – 249, 2013.
Woodall, W. H., R. Koudelik, K. L. Tsui, S., B. Kim, Z. , G. Stoumbos, and C. P. Carvounis, A review and analysis of the Mahalanobis-Taguchi system, *Technometrics*, February, vol. 45, pp. 1-15, 2003.

Biography

Sagar Sikder is a faculty in Statistical Quality Control and Operation Research (SQC & OR) division in Indian Statistical Institute (ISI), India. He earned his B Tech in Mechanical Engineering from REC Warangal, Andra Pradesh, India and ME in Production Engineering from Jadavpur University, Kolkata, India. He is a trainer cum consultant of Six Sigma, SPC & SQC. He has mentored numbers of quality improvement projects in countries leading manufacturing and service sectors. His research interest is in multistage, multivariate SPC.

Indrajit Mukherjee did his Ph.D from the Department of Industrial Engineering and Management, IIT Kharagpur, India. He completed his Master's degree in Quality, Reliability, and Operations Research from the Indian Statistical Institute, Kolkata, India, and his graduation is in Mechanical Engineering from Jalpaiguri Government Engineering College, West Bengal, India, He is working as Assistant Professor in Shailesh J. Mehta School of Management, IIT Bombay, His research work has appeared in many pear reviewed international journals. His research interests are in the domain of multiple response optimization, statistical process control, quality engineering and management, and modelling supply chain variability.

Subhash Chandra Panja is presently working as Associate Professor in the Department of Mechanical Engineering, Jadavpur University, Kolkata, India. He completed his Bachelor in Mechanical Engineering from Jadavpur University, Kolkata. He did Master of Technology and PhD from Industrial Engineering and Management, Indian Institute of Technology Kharagpur. His current areas of research are reliability engineering, process control and improvement, and modelling fuzzy systems. He published a number of research papers both in International Journals and Conferences.

Sasadhar Bera did his Ph.D. degree from Shailesh J Mehta School of Management, IIT Bombay, India, master's in quality, reliability, and operations research from Indian Statistical Institute, Kollkata, India and bachelor's degree in Mechanical Engineering from NIT Durgapur, West Bengal, India. He is presently working as an Assistant Professor in Indian Institute of Management Ranchi, Jharkhand, India. His research work appeared in many pear reviewed international journals and conference proceedings. His research interests are in the area of quality engineering, multiple response optimization, and statistical process control.