# Optimization of Correlated Multiple Response Surfaces with Stochastic Covariate

Taha Hossein Hejazi, Ali Salmasnia, and Mahdi Bastan

*Abstract*—**Optimizing quality engineering problems has become a common problem when there are several correlated product quality characteristics. Moreover, in design of experiments, controlling covariate effects could reduce error and uncovered variances as well as give more insight about the process. This work identifies process variables to analyze correlated multiple responses with stochastic covariates. It also considers dispersion effects and specification limits besides location effects in an integrated framework based on desirability function. At the end, efficacies of the proposed approach are assessed by a numerical example.**

*Index Terms*—**Design of experiments, multiresponse optimization, covariate effects, principal component analysis, and desirability function.**

## I. INTRODUCTION

A common problem in product or process design is selection of design variable settings to optimize different outputs, which are often highly correlated. Several studies have presented approaches addressing multiple quality characteristics but few published papers have focused primarily on the existence of correlation. If correlations among quality characteristics are ignored, engineering designer may miss finding design variable settings which simultaneously improved the quality of all the responses which could lead to an unrealistic solution. Principal Component Analysis can consider the correlation among multiple quality characteristics to obtain uncorrelated components. These components are then substituted into multiple original responses. To optimize several quality characteristics simultaneously, Su and Tong [1] and Antony [2] have tried to combine Taguchi method with principal component analysis. These approaches used the Kaiser's study [3] to select the components whose eigenvalue is bigger than 1 to replace the original responses for further analysis. There are two obvious limitations in Su and Tong [1] and Antony's [2] approaches. First, when the number of principal components whose eigenvalues are greater than 1 is more than one, the required trade off among them is unknown. Second, the total variance in the data does not count in the optimization process of the mentioned methods.

In order to overcome these two main shortcomings in these methods, Liao [4] and Wu and Chyu [5] presented methods based on weighted principal components (WPC) as in the former, PCA is conducted on signal to noise ratio values, whereas in the latter, PCA is conducted on a proportion of quality loss reduction values. In WPC method, all components are considered in order to completely explain variation in all responses. The WPC method uses the coefficient of determination, the proportion of eigenvalue to the number of original responses, as the weight to combine all principal components in order to form a multi-response performance index. Then decide the optimal factor levels by larger-the-better factor effects and smaller-the-better effects in Liao and Wu methods, respectively. Fung and Kang [6] also suggested a WPC method on the bases of Taguchi method. However, in this approach like Su and Tong [1] and Antony [2] only the components whose eigenvalues are greater than 1 are used. Routara *et al.* [7] perform PCA on normalized response variables and compute the Multiresponse Performance Index (MPI) as a weighted summation of PC values. Since the MPI values are not necessarily positive, in the next step SN rations are obtained based on absolute deviations of MPI values from the Ideal point (defined as Combined Quality Loss). Datta et.al [8] applied a combination of Taguchi robust optimization method and PCA to solve correlated MRS problem in which the overall objective function is defined as composite principal components. Since the composite objective is formed by Euclidean norm of principal components values, the negative values of principal components are taken into consider as same as the positive ones. Consequently, this approach could lead into the unreliable results. Ribeiro *et al.* [9] presented another approach for simultaneous optimization of correlated multiple quality characteristics and used score vector of the first principal component obtained from PCA on responses to find optimal design variable setting via response surface methodology (RSM). RSM was used to identify an optimal factor combination that reflects a compromise between the partially incommensurable behavior quality characteristics. As mentioned before, principal components are linear combinations of original response variables. Therefore, when PCA is conducted on signal to noise ratio or proportion of quality loss reduction values, the optimization directions of them might be lost. Regardless of this issue, aforementioned techniques maximize the component values. In other words, they do not correctly consider location effect. To overcome this shortcoming, Tong *et al.* [10] and Wang [11] determined the optimization direction of each component based on the corresponding variation mode chart nevertheless the responses with smaller relative importance could be disregarded. Furthermore, Tong *et al.* [10] and

T. H. Hejazi is with the Department of Industrial Engineering and Management Systems ,Amirkabir University of Technology ,Tehran, Iran (e-mail: t.h.hejazi@aut.ac.ir).

A. Salmasnia is with the Department of Industrial Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: ali.salmasnia@modares.ac.ir).

M. Bastan is with the Department of Industrial Engineering, Faculty of Engineering, Eyvanakey University, Semnan, Iran (e-mail: mbastan@eyc.ac.ir).

Wang [11] used TOPSIS and Relative Closeness to the Ideal Solution (RCIS) indexes as an overall performance index to optimize the multiple quality characteristic problems and to gain the best factor level combination, respectively. Another disadvantage of the PCA-based approaches is that they do not ensure all quality characteristics fall within their specification limits. In this regard, Bashiri and Salmasnia [12] suggested a novel interactive multi response optimization method based on Geoffrion-Dyer-Feinberg (GDF), PCA and conventional desirability function in which two mentioned shortcoming were considered. Hejazi *et al.* [13] represented a novel method based on goal programming to find the best combination of factors so as to optimize multiresponse-multicovariate surfaces with consideration of location and dispersion effects. Also they considered covariate probable values as an objective function which should be maximized. This study aims to present a new general framework which considers all of location and dispersion effects, correlation among the responses and to include specification limits of the responses. In addition, probabilistic covariate is included into the multi-response model to reduce error terms and uncovered variance.

## II. METHODOLOGY

The proposed procedure consists of four phases: Design, PCA, Model building and Optimization. In the first phase, by employing a proper experimental design the significant factors and covariate are identified. Therefore, the required data are gathered. The PCA phase checks the dependency among the output variables. This phase aims to get new independent variables from the original ones. Model building phase consists of empirical modeling of means, variances and principal components with respect to factors and covariate. In the optimization phase, the desirability functions of principal components are constructed. Therefore, an optimization model is used to maximize overall desirability subject to a desired region of all responses. In the rest of this section, parameters and variables are defined, and then the proposed method is described in details.

Parameter and variable definition

The parameters and variables used in the proposed method are defined as follows:

$x$: Design vector
$c$: Covariate
$y_{j\mu}^{min}$: Desired minimum value of $j^{th}$ response.
$y_{j\mu}^{max}$: Desired maximum value of $j^{th}$ response.
$T_{j\sigma^2}$: Target value for variance of the $j^{th}$ response.
$y_{j\sigma^2}^{max}$: Desired maximum value for variance of the $j^{th}$ response.
$f(C)$: Probability distribution function of the covariate.
$\hat{y}_{j\mu}(x)$: Fitted response surface for mean of $j^{th}$ response.
$\hat{y}_{j\sigma^2}(x)$: Fitted response surface for variance of $j^{th}$ response.
$PC_l$: $l^{th}$ Principal component.
$\widehat{PC}_l$: Fitted response surface of $l^{th}$ principal component.
$(L, U)$: Experimental region.
$\Omega$: Feasible region of covariate.

Phase 1- Design
*Step 0*: *identify the significant input variables*

It is important that input variables which influence the response(s) of interest would be identified by a brainstorming session.

*Step 1*: *Select a proper design of experiment*

Depending on the number and the nature of the factors and covariates, a proper design should be select for further analysis.

Phase 2- PCA
*Step 2*: Calculate the sample mean and the sample variance of the $j^{th}$ response under the $i^{th}$ experimental run.

*Step 3*: Calculate the normalized sample mean and variance of the $j^{th}$ response under $i^{th}$ experimental run and normalized probability value of the covariate under the $i^{th}$ run.

*Step 4*: Conduct PCA on the normalized sample mean response, the sample variance response and the covariate probability value to obtain a set of uncorrelated components which are linear combinations of the original responses. The number of retained principal components can be selected based on Kaiser's criterion.

Phase 3- Model building
*Step 6*: Determine probability distribution function of the covariate. Probability distribution function can be defined as follow:

$$\begin{cases} F(c+a) - F(c-a) \leftrightarrow \text{Contineous distribution function for } C \\ P(x = c) \leftrightarrow \text{Discrete distribution function for } C \end{cases}$$

Note that $a$ is an adjusting parameter that helps us finding an interval for calculating probability value of continuous random variables.

*Step 7*: Find the fitted response surfaces of the mean and variance of responses, and the selected principal components.

Phase 4: Optimization
*Step 8*: Obtain the target value of the normalized variance of the $j^{th}$ response that minimizes the fitted response within the experimental region. Request the decision-maker to provide you with an acceptable upper bound.

*Step 9*: Find the target value, the lower and the upper bounds for the desired region of the normalized mean responses.

*Step 10*: Calculate the target value of the normalized probability of the covariate that maximizes probability distribution function. Request the DM to provide you with an acceptable lower bound.

*Step 11*: Obtain the target value, the lower and the upper bounds for the desired region of principal components. Then, construct the conventional desirability function.

*Step 12*: Find the optimal solution $(x^*)$ by solving the optimization model in (1). This model aims to identify $x^*$ that maximizes the overall desirability of the selected principal components within the desired region of mean and variance of original responses.

$$maximize\ D = [\prod_l d(PC_l)] \qquad (1)$$
$$s.t.\ \ y_{j\mu}^{min} \leq \hat{y}_{j\mu}(x) \leq y_{j\mu}^{max}$$
$$T_{j\sigma^2} \leq \hat{y}_{j\sigma^2}(x) \leq y_{j\sigma^2}^{max}$$
$$f(C) \geq \alpha$$
$$X \in (L, U)$$
$$C \in \Omega$$

### III. NUMERICAL EXAMPLE

In this section, a case from the literature is studied to illustrate the applicability of the proposed approach. The case contains two factors and one normally distributed covariate that might affect two responses (TABLE I). The data are generated from three replicates of a Central Composite Design (CCD) and the proposed method is applied to these data.

TABLE I: EXPERIMENTAL RESULT DATA.

| Factors | | Covariate | Response variables | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Expected value | | Estimated Variance | | Probability |
| $x_1$ | $x_2$ | $C \sim N(15, 4^2)$ | $\hat{\mu}_1$ | $\hat{\mu}_2$ | $\hat{\sigma}_1^2$ | $\hat{\sigma}_2^2$ | $f(c)$ |
| -1 | -1 | 19.589 | 35 | 8.65 | 3.51 | 0.79 | 0.0516 |
| 1 | -1 | 17.385 | 30 | 2.88 | 1.16 | 1.70 | 0.0835 |
| -1 | 1 | 22.156 | 29 | 7.95 | 3.57 | 1.98 | 0.0201 |
| 1 | 1 | 9.569 | 43 | 1.97 | 2.75 | 0.75 | 0.0397 |
| -1.414 | 0 | 20.328 | 40 | 5.61 | 3.99 | 1.47 | 0.0411 |
| 1.414 | 0 | 17.519 | 34 | 3.48 | 1.50 | 0.88 | 0.0818 |
| 0 | -1.414 | 19.712 | 29 | 8.31 | 2.33 | 1.66 | 0.0498 |
| 0 | 1.414 | 19.484 | 39 | 8.82 | 3.89 | 1.78 | 0.0532 |
| 0 | 0 | 11.995 | 39 | 5.50 | 1.65 | 1.00 | 0.0752 |
| Desired region | (Min, Target, Max) | | (34 ,36, 38) | (5.75 ,6.25, 6.75) | (0, 0,1.7) | (0, 0,0.8) | (0.17 ,0.197, 0.197) |
| Standard deviation | | | 5.22 | 2.664 | 1.0917 | 0.477 | 0.0213 |

The response surfaces of output variables (mean, variance and first three components) are fitted to input variables (factors and covariate) using MINITAB 15 software. The results are as follows:

$$\widehat{PC}_1 = 6.4483 - 2.1731x_1 - 4.2682x_2 - 0.4163c$$
$$+ 1.1466x_1^2 + 0.2986x_2^2 + 0.1644x_1x_2$$
$$+ 0.1489x_1c + 0.2042x_2c$$

$$\widehat{PC}_2 = -3.9049 + 5.8034x_1 - 2.6785x_2 + 0.304c$$
$$- 1.3848x_1^2 - 0.7588x_2^2 - 0.543x_1x_2$$
$$- 0.2625x_1c + 0.106x_2c$$

$$\hat{y}_{1\mu} = 68.24 - 57.85x_1 + 23.75x_2 - 2.44\,c + 10.24x_1^2$$
$$+ 6.68x_2^2 - 1.73x_1x_2 + 2.82x_1c$$
$$- 1.04x_2c$$

$$\hat{y}_{2\mu} = -5.32 - 5.85x_1 - 1.79x_2 - 0.9\,c - 3.9x_1^2 - 1.89x_2^2$$
$$+ 2.28x_1x_2 - 0.3x_1c + 0.1x_2c$$

$$\hat{y}_{1\sigma^2} = -2.24 - 3.63x_1 + 1.75x_2 - 0.06\,c + 0.91x_1^2$$
$$+ 0.97x_2^2 + 0.17x_1x_2 + 0.14x_1c$$
$$- 0.06x_2c$$

$$\hat{y}_{2\sigma^2} = 0.97 - 2.24x_1 + 2.47x_2 - 0.003\,c + 0.18x_1^2 +$$
$$0.34x_2^2 - 0.84x_1x_2 + 0.11x_1c - 0.12x_2c$$

The last step of the proposed method is to establish the mathematical model and to find best factor setting. In this regard, it is necessary to find the targets of the *PCs*.

Mathematical model of the optimization problem can be formulated as:

$$max \;\; d(PC_1)d(PC_2)$$
*s. t:*

$$34 \le \hat{y}_{1\mu} \le 38;$$

$$5.75 \le \hat{y}_{2\mu} \le 6.75;$$

$$0 \le \hat{y}_{1\sigma^2} \le 1.7;$$

$$0 \le \hat{y}_{2\sigma^2} \le 0.8;$$

$$-1.414 \le x_1 \le 1.414;$$

$$-1.414 \le x_2 \le 1.414;$$
$$0.17 < \varphi\,(c+1) - \varphi(c-1) < 0.197;$$

The last constraint guarantees that the normally distributed covariate is in the predefined range. The optimal solution resulted by solving the above model is represented in TABLE II and also a comparison study between the proposed method and other approaches is given by Table III.

The results indicate that the proposed method outperforms all the methods with respect to PCs' desirability. It is observed that the other approaches do not guarantee limits of the specification region. The values exceeded from the predefined acceptable region are shown by bold numbers.

TABLE II: FINAL RESULTS OF THE CASE.

| Method | $x_1$ | $x_2$ | $C$ | $\hat{y}_{1\mu}$ | $\hat{y}_{2\mu}$ | $\hat{y}_{1\sigma^2}$ | $\hat{y}_{2\sigma^2}$ | $y_{1f}$ | Overall D |
|---|---|---|---|---|---|---|---|---|---|
| Proposed Method | -0.176 | -0.494 | 13.218 | 36.506 | 6.104 | 1.69 | 0.755 | 0.179 | 1 |

TABLE III: COMPARISON STUDY.

| Method | $d(pc_1)$ | $d(pc_2)$ | $\hat{y}_{1\mu}$ | $\hat{y}_{2\mu}$ | $\hat{y}_{1\sigma^2}$ | $\hat{y}_{2\sigma^2}$ | $y_{1f}$ |
|---|---|---|---|---|---|---|---|
| Hejazi et al.[13] | 0.1 | 0.819 | 36 | 6.25 | 1.51 | 0.88 | 0.172 |
| Derringer [14] | 0.224 | 0.934 | 36 | 6.252 | 1.54 | 0.85 | 0.17 |
| Kim and Lin [15] | 0 | 0.914 | 33.36 | 5.722 | 1.33 | 0.79 | 0.15 |
| Kovach and Cho [16] | 0 | 0.95 | 36 | 6.25 | 1.5 | 0.925 | 0.169 |
| Pignatiello [17] | 0.271 | 0.927 | 35.92 | 6.07 | 1.512 | 0.809 | 0.17 |
| Proposed Method | 1 | 1 | 36.5 | 6.104 | 1.69 | 0.755 | 0.179 |

## IV. CONCLUSION

In this paper, a mathematical model has been developed to find the best settings of variables to optimize multiresponse models where a stochastic covariate affects the responses. In addition, this work studies correlated responses by an approach based on PCA and also considers the dispersion effects. Another advantage of the proposed approach is to model multiresponse problem considering single covariate with known distribution function. Following subjects can be suggested as futures research:

1) Consideration of multiple covariate models.
2) Consideration of qualitative variables.
3) Considering the variances of predicted responses as a separate objective function.

## REFERENCES

[1] C. T. Su and L. I. Tong, "Multi-response robust design by principal component analysis," *Total Quality Management*, vol. 8, pp. 409-416,1997.
[2] J. Antony, "Multi-response optimization in industrial experiments using Taguchi's quality loss function and principal component analysis," *Quality and Reliability Engineering International*, vol. 16, pp. 3-8, 2000.
[3] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educ. Psychol. Meas* vol. 20, pp. 141-151, 1960.
[4] H. C. Liao, "Multi-response optimization using weighted principal component," *International Journal of Advanced Manufacturing Technology*, vol. 27, pp. 720-725, 2006.
[5] F. C. Wu and C. C. Chyu, "Optimization of correlated multiple quality characteristics robust design using principal component analysis," *Journal of Manufacturing Systems*, vol. 2, pp. 134-143, 2004.
[6] C. P. Fung and P. C. Kang, "Multi-response optimization in friction properties of PBT composites using Taguchi method and principal component analysis," *Journal of Materials Processing Technology*, vol. 170, pp. 602-610, 2005.
[7] B. C. Routara, S. D. Mohanty, S. Datta, A. Bandyopadhyay, and S. S. Mahapatra, "Combined quality loss (CQL) concept in WPCA-based Taguchi philosophy for optimization of multiple surface quality characteristics of UNS C34000 brass in cylindrical grinding,"

*International Journal of Advanced Manufacturing Technology*, DOI 10.1007/s00170-010-2599-1, 2010.
[8] S. Datta, G. Nandi, A. Bandyopadhyay, and P. K. Pal, "Application of PCA-based hybrid Taguchi method for correlated multicriteria optimization of submerged arc weld: a case study," *International Journal of Advanced Manufacturing Technology*, vol. 45, pp. 276-286, 2009.
[9] J. S. Ribeiro, R. F. Teofilo, F. Augusto, and M. M. C. Ferreira, "Simultaneous optimization of the microextraction of coffee volatiles using response surface methodology and principal component analysis," *Chemometrics and Intelligent Laboratory Systems* vol. 102, pp. 45-52, 2010.
[10] L. I. Tong, C. H. Wang, and H. C. Chen, "Optimization of multiple responses using principal component analysis and technique for order preference by similarity to ideal solution," *International Journal of Advanced Manufacturing Technology*, vol. 27, pp. 407-414, 2005.
[11] C. H. Wang, "Dynamic multi-response optimization using principal component analysis and multiple criteria evaluation of the grey relation model," *International Journal of Advanced Manufacturing Technology*, vol. 32, pp. 617-624, 2007.
[12] M. Bashiri and A. Salmasnia, "Decision making for interactive optimization of correlated desirability function," *Proceeding of IEEE International conference on Industrial Engineering and Engineering Management,* pp. 2075-2079, 2009.
[13] T. H. Hejazi, M. Bashiri, K. Noghondarian, and A. C. Atkinson, "Multiresponse optimization with consideration of probabilistic covariates," *Journal of Quality and Reliability Engineering International*. DOI: 10.1002/qre.1133, 2010.
[14] G. Derringer and R. Suich, "Simultaneous optimization of several response variables," *Journal of Quality Technology*, vol. 12, pp. 214-219, 1980.
[15] K. J. Kim and D. K. J. Lin, "Simultaneous Optimization of mechanical properties of steel by polynomial regression functions," *Applied Statistics*, vol. 49, pp. 311-325, 2000.
[16] J. Kovach and B. R Cho, "Development of a multidisciplinary-multiresponse robust design optimization model," *Engineering Optimization*, vol. 40, pp. 805 - 819, 2008.
[17] J. Pignatiello, "Strategies for robust multiresponse quality engineering," *IIE Transactions,* vol. 25, pp. 5-15, 1993.

**Taha Hossein Hejazi** was born in mashhad, 1985. He received a MSc in Industrial Engineering at Shahed University of Tehran, Iran. He is PhD candidate at Amirkabir University of Technology, Tehran, Iran. His primary research interests include applied and engineering statistics especially quality engineering and simulation.

**Ali Salmasnia** is a doctoral student in the Department of Industrial Engineering at the Tarbiat Modares University in Tehran, Iran. He received his B.S. degree from the Mazandaran University of Science and Technology in Babol, Iran, and his M.S. degrees in industrial engineering form the Shahed University in Tehran, Iran. His current research interests include Multiple Response Optimization, Reliability and Project Management. He

has published in Applied Soft Computing, Neurocomputing, IEEE Transactions on Engineering Management, International Journal of Advanced Manufacturing Technology, International Journal of Applied Decision Sciences, International Journal of Industrial Engineering & Production Management, Australian Journal of Basic and Applied Sciences, and Advanced Materials Research.'

**Mahdi Bastan** was born in 1 may 1985 in Tehran, Iran. Live in Tehran City, Iran. He Works in the Eyvanakey University,Semnan, Iran. He graduated from IHU, Theran, Iran in Socio-Economic Systems Engineering at 2011. His main research interests are System dynamics, Quality Engineering, Entrepreneurship and Strategic planning, he has membership in the below institutions: American Society for Quality. System Dynamics Society. Iranian Operations Research Society. Iranian institute of industrial Engineering. Iranian institute of Strategic Management.