

A desirability functions-based approach for simultaneous optimization of quantitative and ordinal response variables in industrial processes

^{1*}Surajit Pal, ²Susanta Kumar Gauri

¹*SQC & OR Unit, Indian Statistical Institute, 110, N. M. Road, Chennai 600 029, INDIA*

²*SQC & OR Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700 108, INDIA*

**Corresponding Author: e-mail: surajitpal@hotmail.com*

Abstract

The most important step for producing high quality products is the optimum utilization of the manufacturing processes and its resources, which can be accomplished by using optimal process settings. Extensive research works are reported in literature for optimization of process settings with respect to single as well as multiple quantitative response variables. However, in real world, often some aspects of product quality (e.g. dimensions, yield etc.) are measured quantitatively and some other aspects of product quality (e.g. color, finish etc.) are assessed qualitatively like 'poor', 'average', 'good' etc., which can be treated as ordered categorical or ordinal variables. Only a few researchers have attempted to tackle the problem of simultaneous optimization of quantitative and ordinal response variables. But none of these approaches results in an efficient optimal solution with respect to the desirability values of all the individual response variables. It motivated us to develop a more useful approach for simultaneously optimizing industrial processes with respect to quantitative as well as ordinal response variables. In the proposed method, the concept of desirability functions for continuous quantitative variables is extended for ordinal response variables, and then process settings is optimized considering geometric mean of the individual desirability functions of continuous as well as ordinal response variables. A set of experimental data, which have been analyzed by the past researchers, are analyzed using the proposed method and the related results are compared. The comparison of results reveals that the proposed method leads to better optimal solution than the other methods.

Keywords: Quantitative responses, Ordinal responses, Multiple regression, Ordinal logistic regression, Desirability function, Optimization

DOI: <http://dx.doi.org/10.4314/ijest.v10i1.6>

1. Introduction

In order to survive in today's highly competitive world, it is essential for a manufacturing organization to produce high quality products consistently with minimum manufacturing cost. Most of the products are complex in nature and quality of such a product is often described by several quality characteristics – some of them are quantitative and some of them are qualitative. In order to improve product quality, all these quantitative and qualitative responses must be optimized simultaneously. The products' quality characteristics (response variables) are generally affected by the setting combination of several input variables of the manufacturing process. Therefore, it is essential to choose the best setting combination of input process variables so that all response variables are optimized.

Over the last three decades, Taguchi's robust design methodology has been widely employed in industries for optimizing input process parameters. Taguchi (1986) employed an orthogonal array to conduct experiments, which facilitates assessing

the effects of several input process variables using lesser number of experiments. Taguchi applied a quality loss function to evaluate product quality and employed the signal-to-noise ratio (SNR) with simultaneous consideration of achieving the target and reducing variability around the target value of the response variable. However, Taguchi (1986) focused on optimization of a single quality characteristic only, whereas most of the modern manufacturing processes usually have several response variables.

Aiming to satisfy the need of the industries, several methodologies were developed utilizing classical statistical techniques like response surface methodology and various optimization search algorithms, e.g. Desirability function-based approach (Derringer and Suich, 1980), Mahalanobis distance-based approach (Khuri and Conlon, 1981), multiple regression and linear programming-based approaches (Vining and Myers, 1990; Castillo and Montgomery, 1993; Pignatiello and Joseph, 1993) and utility function-based approach (Kumar et al., 2000). Among various approaches for the multi-response optimization, the desirability function-based approach (Derringer and Suich, 1980) has gained the maximum popularity in solving the multi-response optimization problems. Subsequently, many researchers (Castillo et al., 1996; Kim and Lin, 2000; Jeong and Kim, 2003; Wu, 2005; Kushwaha et al., 2013; Candiotti et al., 2014; Salmansia and Bashiri, 2015; Ahmad et al., 2017) modified the desirability functions and used it for optimization of multiple quantitative response variables. All these approaches are mathematically rigorous and consequently, are not easy to implement in industries. Several other researchers, therefore, have taken interest in developing appropriate methodologies for optimizing multiple responses under Taguchi's robust design framework (Tong and Hsieh 2000; Chiang and Su 2003; Tong et al. 2005; Pan et al. 2007; Pal and Gauri, 2010; Khanna et al., 2015; Priyadarshini et al., 2015; Fard, et al., 2016; Ghani et al., 2017). However, all these researchers make an implicit assumption that all the responses are quantitative variables and thus, their proposed methodologies for multi-response optimization are applicable only if all the response variables are quantitative in nature.

But, in reality, some quality characteristics of a product are often expressed as ordered categorical variables (ordinal variables). Also, owing to the inherent nature of the quality characteristic or the convenience of the measurement technique and cost-effectiveness, few aspects of quality of products are described in terms of ordered categorical data, i.e. ordinal data. For analyzing such ordinal data, Taguchi (1986) proposed accumulation analysis (AA) method. Subsequently several authors (Agresti, 1986; Box and Jones, 1986; Hamada and Wu, 1986; Nair, 1986; Koch et al., 1990; Chipman and Hamada, 1996) critically investigated Taguchi's AA method and highlighted the weaknesses in Taguchi's AA method. As described by them, the main weakness of Taguchi's AA method is that although it has reasonable power for detecting location effects, it performs poorly in identifying the dispersion effect. Few researchers have proposed some alternative approaches for optimizing an ordinal response, e.g. scoring schemes approach (Nair, 1986), Bayesian generalized linear modelling (BA-GLM) approach (Chipman and Hamada, 1996), weighted probability scoring scheme (WPSS) approach (Jeng and Guo, 1996), weighted signal-to-noise ratio (WSNR) approach (Wu and Yeh, 2006), expected loss minimization approach (Asiabar and Ghomi, 2006) and heuristic/meta-heuristic approach (Bashiri et al., 2012). Again, all these methods deal with a single ordinal response.

However, in real world, there are many products for which the overall quality is represented partly by some quantitative response variables and partly by some ordinal response variables. In such cases, simultaneous optimization of both types of response variables is essential for improving the quality performance. Therefore, availability of an appropriate methodology which is capable of optimizing all the quantitative and ordinal variables simultaneously can truly satisfy the practical need of the industries. It is observed that only Hsieh and Tong (2001), Wu (2008) and Liao et al. (2014) have attempted to optimize quantitative and ordinal response variables simultaneously under Taguchi's framework of robust design approach. Hsieh and Tong (2001) applied artificial neural networks for simultaneous optimization of quantitative and ordinal responses. Liao et al. (2014) developed an ant colony optimization (ACO) algorithm that can tackle mixed-variable optimization problems. On the other hand, Wu (2008) treated the ordinal data as continuous variables assuming weight for each category is proportional to the quality loss. Then, he computed total quality loss and transformed it to signal-to-noise ratio (SNR), which is maximized to derive the optimal process settings. The problem with Wu's (2008) approach is that it does not lead to efficient optimal solution. On the other hand, the problem with the artificial intelligence-based techniques or ACO algorithm is that these do not help engineers to learn efficient engineering experiences during the period of the optimizing process. Also, these are quite difficult to be employed in industries. In this article, a new procedure for simultaneous optimization of quantitative and ordinal response variables is proposed that can be easily implemented. In the proposed method, a new performance metric is defined that integrates logistic regression technique, multiple regression technique and desirability function concept. The process parameter is optimized by maximizing the overall desirability index.

This article is organized as follows. Section 2 reviews literature related to optimization of multiple quantitative response variables and ordinal response variables. Section 3 describes the framework of the proposed method for simultaneous optimization of quantitative and ordinal response variables. Section 4 summarizes the necessary steps for implementation of the proposed method. Analysis of a set of experimental data taken from literature, related results and comparison of optimization performance are presented in section 5. The article is concluded in section 6.

2. Literature Review

Most often, quality characteristics of a manufactured product are expressed as quantitative response variables, and therefore, traditionally the problem of multi-response optimization is focused on the assumption that all the response variables are quantitative (particularly, continuous) in nature. Derringer and Suich (1980) first proposed the use of a desirability function, which was originally introduced by Harrington (1965), as a metric for optimization of multiple response variables. Subsequently, many researchers (Castillo et al., 1996; Kim and Lin, 2000; Jeong and Kim, 2003; Wu, 2005; Kushwaha et al., 2013; Candioti et al., 2014; Salmasia and Bashiri, 2015; Ahmad et al., 2017) modified the desirability functions and used it for optimization of multiple quantitative response variables. Till date, the desirability function-based approaches are most commonly used for optimizing multiple quantitative response variables. However, there are several other approaches reported in literature for optimizing multiple quantitative response variables, e.g. Mahalanobis distance-based approach (Khuri and Conlon, 1981), multiple regression and linear programming-based approaches (Vining and Myers, 1990; Castillo and Montgomery, 1993; Pignatiello and Joseph, 1993), utility function-based approach (Kumar et al., 2000). Some of these mathematically rigorous techniques aim at minimizing deviation from target as well as maximizing robustness to noise, which conform to Taguchi's philosophy. However, those are not easily comprehensible and thus, are quite difficult to implement in industry. Keeping this in mind, a group of researchers (Tong and Hsieh 2000; Chiang and Su 2003; Tong et al. 2005; Pan et al. 2007; Pal and Gauri, 2010; Khanna et al., 2015; Priyadarshini et al., 2015; Fard, et al., 2016; Ghani et al., 2017) have proposed some simplified approaches for the multi-response optimization, which can be applied under the framework of Taguchi method (1986) of experimentation and analysis. In these approaches, the quality loss or SNR of individual responses is converted first into an overall process performance index (PPI) and then, the optimal settings of the input variables are determined by examining the level averages on the PPI.

The importance of analyzing ordered categorical quality characteristics (ordinal data) also got attention of several researchers. Taguchi (1986) primarily developed the accumulation analysis (AA) method to optimize input factor settings for an ordinal response variable. In this method, the number of observations indifferent categories is first converted to number of observations by cumulative categories and then, probabilities for the cumulative categories are computed. The optimal input factor settings are determined for the desired cumulative category. Taguchi (1986) recommended the use of Omega (Ω) transformation to predict the defect counts from the probabilities of the cumulative categories. Subsequently several authors (Agresti, 1986; Box and Jones, 1986; Hamada and Wu, 1986; Nair, 1986; Koch et al., 1990; Chipman and Hamada, 1996) critically investigated Taguchi's AA method and highlighted the weaknesses in Taguchi's AA method. As described by them, the main weakness of Taguchi's AA method is that although it has reasonable power for detecting location effects, it performs poorly in identifying the dispersion effect. Subsequently, many researchers proposed some alternative approaches for optimizing an ordinal response. Nair (1986) presented two scoring schemes to identify the location and dispersion effects. These two sets of scores are used to find the contribution of factor levels on location and dispersion effects and determining the optimal factor setting combination. McCullagh (1986) developed a logistic regression model using maximum likelihood method to assess location and dispersion effects of ordered categorical data. Jeng and Guo (1996) proposed a weighted probability scoring scheme (WPSS). They suggested using mean square deviation (MSD) as a performance measure to avoid making compromise between location and dispersion effects. The optimal factor setting is obtained by minimizing MSD. Chipman and Hamada (1996) proposed a Bayesian model for the ordinal data. They employed the combination of a generalized linear model with Bayesian estimation techniques to optimize the factor settings. Asiabar and Ghomi (2006) proposed a simple method for optimization of ordered categorical data using expected loss minimization. Wu and Yeh (2006) presented a weighted signal-to-noise ratio (WSNR) method for analyzing ordered categorical data, and compared optimization performance of four methods: AA method, WSNR method, Scoring Scheme and WPSS methods. They observed that efficiency of WSNR method is inferior to Scoring Scheme method but better than the other two methods. Bashiri et al. (2012) optimized an ordinal response variable by using ordinal logistic regression and heuristic/meta-heuristic optimization algorithms.

However, analyses of quantitative response variables and ordinal response variables separately usually are not much useful. This is because, most often, the optimal parametric settings with respect to quantitative response variables and ordinal response variables are different. Ideally, process settings have to be optimized taking into consideration the quantitative as well as ordinal response variables simultaneously. The problems of simultaneous handling of quantitative and ordinal data have got least attention of the researchers. Only Hsieh and Tong (2001), Wu (2008) and Liao et al. (2014) have attempted to optimize simultaneously the quantitative and ordinal responses. Hsieh and Tong (2001) applied artificial neural networks for simultaneous optimization of quantitative and ordinal responses. Liao et al. (2014) developed an ant colony optimization (ACO) algorithm that can tackle mixed-variable optimization problems. On the other hand, Wu (2008) treated the ordinal data as continuous variables assuming weight for each category is proportional to the quality loss. Then he computed total quality loss and transformed it to signal-to-noise ratio (SNR), which is maximized to derive the optimal process setting. The problem with the artificial intelligence-based techniques (Hsieh and Tong, 2001) or ACO algorithm is that these do not help

engineers to learn efficient engineering experiences during the period of the optimizing process, whereas, Wu's (2008) approach does not lead to efficient optimal solution. So there is a need for developing a better method for simultaneous optimization of quantitative and ordinal response variables.

3. The Framework of the Proposed Method

One of the most popular performance metrics for simultaneous optimization of multiple continuous response variables is the desirability function (Derringer and Suich, 1980). In this approach, the individual responses are estimated first through regression techniques, and then, each estimated response is transformed into a desirability value. The individual desirabilities are then combined using the geometric mean to get the overall desirability, i.e. the performance metric. Our basic idea was that if it is possible to obtain desirability functions for the ordinal response variables, then the geometric mean of individual desirability indices of continuous as well as ordinal response variables can be considered as the performance metric for simultaneous optimization of quantitative and ordinal response variables. It is well established that the probabilities in each category of the ordinal variable can be predicted by using the ordinal logistic regression model (Agresti, 2010). Jeng and Guo (1996) proposed a weighted probability scoring scheme (WPSS) for estimating the location and dispersion effects (scores) for an ordinal variable using the predicted probabilities. We observed that both location score (LS) and dispersion score (DS) are quantitative, and therefore, considering LS and DS as LTB and STB type variables, the desirability functions for location effect and dispersion effect of an ordinal response variable can be defined appropriately without loss of generality. This indicates that our idea for simultaneous optimization of quantitative and ordinal response variables using overall desirability index is feasible.

According to our proposed method, optimization of a process having quantitative as well as ordinal response variables would require the following four steps: (i) obtaining individual desirability index for each quantitative response variable, (ii) obtaining desirability indices for location effect and dispersion effect of each ordinal response variable, (iii) obtaining overall desirability index and (iv) optimizing the overall desirability index.

Let us assume a process for describing the proposed procedure. Suppose the process has m control factors (X_1, X_2, \dots, X_m), which are to be optimized with respect to r continuous response variables and s ordinal response variables (each ordinal response have k levels or categories of ordered categorical data). Suppose, an orthogonal experimentation with n trial runs is carried out and each trial run is replicated q times. In each replication, values of all the response variables are observed.

3.1 Obtaining Desirability Function for Quantitative Response Variables

For obtaining the desirability function for a quantitative response variable, the first requirement is to establish the most appropriate regression equation for prediction of the response variable. In an experimentation set up, different levels of the control factors in a trial run are represented by the numerical values like 1, 2, 3 etc., i.e. each trial run has different set level values of the input variables \mathbf{X} . Using these set level values of the input variables \mathbf{X} and the response values of Y_i corresponding to n trial runs, appropriate multiple regression equation for prediction of expected mean of i^{th} quantitative response variable (Y_i) ($i = 1, 2, 3, \dots, r$) can be worked out. The regression model will be as given in Eqn. (1).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \beta_{m+1} X_1^2 + \dots + \beta_p X_{m-1} X_m + V \quad (1)$$

where β_i 's are regression coefficients, m is the number of input variables, p is the total number of regressor terms in respective equations and V is random error. Generally, it is assumed that V is independently and normally distributed with mean value as zero and with constant variance σ^2 . It is important to note that the number of regressor terms in any equation should be smaller than the number of trial runs in the designed experiment.

The regression coefficients can be obtained by the least square method. The option of performing multiple linear regression analysis is available in Microsoft Excel and in many statistical software packages, e.g. MINITAB and STATISTICA. Using these packages, appropriate multiple regression equations can be established. However, diagnostic checks must be performed for validating each fitted regression equations. The adequacy of fitted model can be checked using ANOVA and an F -test for significance of regression. Any kind of possible anomalies can be detected by examining residual analysis in terms of various plots, e.g. normality plot of residuals, plot of residual versus individual regression variable etc. If, after the diagnostic checks, no serious violations of model assumptions are detected, then the regression equations are assumed to be adequate fit to predict means of the response variable.

The concept of desirability function was originally proposed by Harrington (1965) for continuous response variable. It is a function by which an estimated response variable \hat{Y}_i is transformed into desirability value d_i , where $0 \leq d_i \leq 1$. The value of d_i increases as the desirability of the response increases. If the product characteristic is in an unacceptable range, the desirability value is 0 and if the product characteristic is at the optimum value, the desirability value is 1.

Derringer and Suich (1980) modified this desirability function for optimization of multiple response variables. They defined three classes of desirability functions for nominal-the-better (NTB), smaller-the-better (STB) and larger-the-better (LTB) types of response variables. For the NTB type, the desirability function is defined as

$$d = \begin{cases} \left| \frac{y - LSL}{T - LSL} \right|^s, & LSL \leq y \leq T \\ \left| \frac{y - USL}{T - USL} \right|^t, & T \leq y \leq USL \\ 0, & y < LSL \text{ or } y > USL \end{cases} \quad (2)$$

where the exponents s and t are the shape constants of the desirability function, and LSL and USL are respectively the lower and upper specification limit for the NTB type response variable with a target value T . In general, the shape constants are chosen in the range from 0.01 to 10. We use both the shape constants are equal to 2 so as to look similar as Taguchi's quadratic loss function.

For the STB type, the desirability function is defined as

$$d = \begin{cases} \left| \frac{y - USL}{a - USL} \right|^t, & a \leq y \leq USL \\ 0, & y > USL. \end{cases} \quad (3)$$

where a is a smallest possible value for response Y .

For the LTB type, the desirability function is defined as

$$d = \begin{cases} \left| \frac{y - LSL}{USL - LSL} \right|^s, & LSL \leq y \leq USL \\ 0, & y < LSL \\ 1, & y > USL \end{cases} \quad (4)$$

3.2 Obtaining Desirability Functions for Location and Dispersion effects of Ordinal Response

Let us assume that any one of s ordinal response variables $Y_l (l=1,2,\dots,s)$ has k number of ordered categories, denoted by $1,2,\dots,k$, where $k \geq 3$, integer. From the experimental data, for each of n experimental runs, the number of observations (frequencies) in each category from q number of repeated samples are first observed. These frequencies can easily be converted into individual and cumulative probabilities for each of k categories. Using ordinal logistic regression, these cumulative probabilities are modeled as a function of input controllable variables. Generally, a logit transformation of cumulative probabilities are used as a link function while modeling the cumulative probabilities as a function of input variables and accordingly, this type of ordinal regression models are named as cumulative logit models. If there are k categories of the ordinal variable Y_l , then there will be $k-1$ cumulative logit models for describing first $k-1$ cumulative probabilities of Y_l .

Let p_i denotes the probability that the $u^{\text{th}} (u = 1, 2, \dots, q)$ sample of the ordinal response variable Y_l falls in i^{th} category, that is, $p_i = P(y_{lu} = i)$, where $i = 1, 2, \dots, k$. The transformation $y_i = \ln \frac{p_i}{1 - p_i}$ is called the logit transformation of probability p_i .

Logit of cumulative probability up to i^{th} category is defined as

$$\log it[P(Y_{lu} \leq i)] = \log \frac{P(y_{lu} \leq i)}{1 - P(y_{lu} \leq i)}, \quad l = 1, 2, \dots, s; u = 1, 2, \dots, q \quad (5)$$

These logits of cumulative probabilities are modeled as function of input controllable variables as shown below:

$$\log it[P(Y_{lu} \leq i)] = r_i + \sum_{j=1}^m S_j x_j, \quad i = 1, 2, \dots, k-1 \quad (6)$$

All the cumulative logit models for all response categories of the ordinal variable Y_l have the same effect of controllable variables. That means, the S_j coefficients of j^{th} input variable of response Y_l are same in all $k-1$ logit models. The only difference in the models from one response category to the next is the intercepts r_i . The $\{r_i\}$ are increasing in i because cumulative probability $P(y_l = i)$ increases in i and logit transformation is an increasing function of this cumulative probability.

From the cumulative logit models, the cumulative probabilities of i^{th} category of the ordinal variable Y_l can be computed as

$$P(y_l \leq i) = \frac{\exp(r_i + \sum_{j=1}^m S_j x_j)}{1 + \exp(r_i + \sum_{j=1}^m S_j x_j)}, \quad i = 1, 2, \dots, k-1 \quad (7)$$

From the cumulative probabilities of first $k-1$ categories, individual probabilities of all k categories can be easily obtained. Individual probability of i^{th} category can be computed as

$$P(y_l = i) = \frac{\exp(r_i + \sum_{j=1}^m S_j x_j)}{1 + \exp(r_i + \sum_{j=1}^m S_j x_j)} - \frac{\exp(r_{i-1} + \sum_{j=1}^m S_j x_j)}{1 + \exp(r_{i-1} + \sum_{j=1}^m S_j x_j)}, \quad i = 1, 2, \dots, k-1 \quad (8)$$

For the first category, individual and cumulative probabilities are same. For the individual probability of 2^{nd} category, the cumulative probability of 1^{st} category is to be subtracted from the cumulative probability of 2^{nd} category and so on. For the individual probability of the last k^{th} category, the cumulative probability of $(k-1)^{\text{th}}$ category is to be subtracted from 1, because total cumulative probability of k^{th} category is always equal to one. Multiplying individual probabilities by the sample size, the expected individual frequencies of all k categories of the l^{th} ordinal response variable can be computed. Using similar method, the cumulative logit regression models for all s ordinal response variables can be established. The regression coefficients of the ordinal logistic regression models can be obtained by using any statistical package e.g. Minitab, SPSS etc. However, usual goodness-of-fit tests and diagnostic checks must be performed for validating the cumulative logit models for the ordinal data.

For computing location and dispersion effects in case of an ordinal response variable, Jeng and Guo (1996) proposed a weighted probability scoring scheme (WPSS). In this method, the categorical data is listed first in a progressively undesirable order, and then, the more desirable category is given more weight to make the location effect as the larger-the-better (LTB) type quantitative response variable. Let there are K categories of an ordinal quality characteristic with first category representing the best quality and a weight $w_k = K - (k - 1)$, $k = 1, 2, \dots, K$ is assigned to each category. The location score LS and the dispersion score DS for a particular setting combination is then defined as

$$LS = \sum_{k=1}^K w_k p_k, \quad (9)$$

$$DS = \sum_{k=1}^K [w_k p_k - (\text{Target Value})_k]^2 \quad (10)$$

where p_k is the proportion of observation in k^{th} category and the target values are defined as $[K, 0, 0, \dots, 0]$ for categories 1, 2, ..., K , respectively.

It may be noted that both LS and DS are quantitative. The maximum possible value of LS is K and the minimum possible value is 1. Higher value of LS is desirable, and therefore, LS may be considered as LTB type quantitative response variable with $USL = K$ and $LSL = 1$. On the other hand, the maximum possible value of DS is $[K^2 + (K-1)^2]$ and the minimum possible value is 0. Smaller value of DS is desirable, and therefore, DS may be considered as STB type quantitative response variable with $USL = [K^2 + (K-1)^2]$ and $LSL = 0$. Thus, without loss of generality, both LS and DS can be converted into desirability indices using the following desirability function

$$d_{LS} = \left(\frac{LS - 1}{K - 1} \right)^2 \quad (11)$$

$$d_{DS} = \left[\frac{DS - \{K^2 + (K-1)^2\}}{K^2 + (K-1)^2} \right]^2 \quad (12)$$

3.3. Formulation of the Performance Index

For the purpose of simultaneous optimization of quantitative and ordinal response variables, weighted geometric mean of desirability indices of the individual quantitative response variables and desirability indices of LS and DS of all the ordinal response variables may be considered as the overall desirability index. The overall desirability index D , therefore, at a chosen setting combination can be obtained

$$D = \left\{ (d_1^Q \times d_2^Q \times \dots \times d_r^Q) \times (d_1^{LS} \times d_1^{DS} \times d_2^{LS} \times d_2^{DS} \times \dots \times d_s^{LS} \times d_s^{DS}) \right\}^{1/(r+2s)} \quad (13)$$

where d_i^Q is the desirability index of i^{th} ($i = 1, 2, \dots, r$) quantitative response variable at the chosen setting combination, and d_j^{LS} and d_j^{DS} ($j = 1, 2, \dots, s$) are the desirability indices for location score and dispersion score of j^{th} ordinal response variable at the chosen setting combination. As per definition, higher desirability index is better.

3.4 Optimization of the Performance Index

It is important to note that overall desirability index D is a function of some predicted values, which are again functions of the control factor levels. Thus, D is essentially a function of the control factor levels. So the object of optimization is to determine the level values of the control factors (input variables) that maximize the D value.

The task of determining optimal level values of the control factors can be accomplished by changing level values of the input variables and comparing the resulting D values. The 'Solver' tool of Microsoft Excel package is very effective for performing such enumerative search and finding the optimal level values for the input variables. The Solver tool employs the generalized reduced gradient (GRG) method for optimization, proposed by Castillo and Montgomery (1993). While running the 'Solver' tool it is necessary to specify the range of levels for the input variables. In certain cases, where the input variable takes only discrete values, the integer restriction for that input variable need to be specified.

It may be noted that for obtaining optimal solution using the 'Solver' tool one has to choose an arbitrary setting combination (say, existing combination) first, and then to carry out all the necessary computations for prediction of quantitative responses, LS and DS scores of ordinal responses, desirability indices for quantitative responses, LS and DS of ordinal responses and the overall desirability for the chosen combination in Excel worksheet. Finally he/she has to run the 'Solver' tool.

4. The Procedure for Implementation of the Proposed Method

The proposed method for simultaneous optimization of quantitative and ordinal response variables can be implemented using the following five steps:

Step 1: Identify the types of quantitative response variables, i.e. if LTB, STB or NTB type. For each experimental trial, record the replicated values of each quantitative response variable. On the other hand, identify different categories for each ordinal response variable and arrange the categorical data in a progressively undesirable order. Then compute cumulative probabilities for all the categories of an ordinal response variable.

Step 2: Fit appropriate multiple regression equations for prediction of the expected means for all the quantitative response variables using models shown in Eqn. (1). On the other hand, establish appropriate cumulative logit model for prediction of cumulative probability up to a given category of an ordinal response variable using Eqn. (6). For an ordinal response with k categories, there will be $k-1$ cumulative logit models for describing first $k-1$ cumulative probabilities.

Step 3: Choose an arbitrary combination of control factor levels (say, existing combination), and then, obtain the overall desirability index for the chosen setting combination in a Microsoft Excel worksheet as follows:

- (a) Obtain the predicted means of all quantitative response variables using respective regression equations. Similarly, obtain predicted probabilities in each category of each ordinal response variable using respective ordinal logistic regression model.
- (b) Transform the predicted mean of each quantitative response variable into desirability indices using Eqns. (2-4) as appropriate depending on the type of the response variable. Also transform the predicted probabilities of each ordinal response variable into location and dispersion scores, LS and DS, using Eqns. (9-10), and transform these further into desirability indices for location effect and dispersion effect using Eqns. (11-12).
- (c) Combine all the desirability indices of all individual response variables (quantitative as well as ordinal) in a single overall desirability index, D , using Eqn. (13).

Step 4: Using 'Solver' tool of Microsoft Excel package, determine the level values of the input variables (control factors) that maximize the D value.

Step 5: Obtain the expected values (along with confidence intervals) of each quantitative response variable at the optimal setting condition using the relevant multiple regression equation and expected frequencies at each category of the ordinal response variable. Then, carry out the confirmatory trial and verify that the actual results conform to the expected results.

5. Analysis and Results

For illustrating the application of the proposed method and evaluate the effectiveness of the proposed method, the case study data on ion implantation process of a Taiwanese IC fabrication manufacturer, described by Hsieh and Tong (2001), is analyzed. In this case study, two responses were considered: the first one was one was quantitative (on the measurement of ion amount) and the second one was qualitative (total defect counts of 36 sensitive areas in a wafer). The quantitative response is of NTB type with a target value of 1000 and with LSL and USL of 800 and 1200 respectively. The qualitative response is ordinal variable and it includes five categories of defect situation: very good, good, not good and not bad, bad, and very bad and those are denoted by grade I, II, III, IV, and V. Six control factors A – F were studied. Among them, A is discrete with 2 levels and other five factors B – F each having 3 levels. An L_{18} orthogonal array was planned and each experiment was conducted with two repetitions. The experimental data in summarized form is shown in Table 1.

Table 1. Experimental data from ion implantation process

Expt No.	Factors						Ion Amount (IA) (Y_1)		Defect Count (DC) (Y_2)				
	A	B	C	D	E	F	Rep1	Rep2	I	II	III	IV	V
1	1	1	1	1	1	1	745.2	741.4	33	3	0	0	0
2	1	1	2	2	2	2	968.3	972.1	24	5	6	1	0
3	1	1	3	3	3	3	800.2	796.1	6	2	20	8	0
4	1	2	1	1	2	2	795.9	797.8	0	14	4	4	0
5	1	2	2	2	3	3	791.4	796.6	2	12	4	12	16
6	1	2	3	3	1	1	800.4	802.1	4	12	20	4	8
7	1	3	1	2	1	3	912.2	908.2	0	2	6	14	14
8	1	3	2	3	2	1	650.0	645.7	10	2	8	4	12
9	1	3	3	1	3	2	651.2	650.3	0	0	0	24	12
10	2	1	1	3	3	2	1075.1	1072.5	34	0	2	0	0
11	2	1	2	1	1	3	1314.0	1616.1	30	2	4	0	0
12	2	1	3	2	2	1	884.4	890.5	10	10	12	0	4
13	2	2	1	2	3	1	884.3	886.6	14	8	10	4	0
14	2	2	2	3	1	2	817.4	826.5	8	16	12	0	0
15	2	2	3	1	2	3	796.0	800.1	0	8	6	4	18
16	2	3	1	3	2	3	819.8	816.1	18	12	6	0	0
17	2	3	2	1	3	1	821.8	824.2	10	6	0	4	16
18	2	3	3	2	1	2	732.4	735.6	0	4	2	6	24

As described in step 2, the multiple regression equation for prediction of IA (quantitative response variable) is now established as a function of controllable variables A-F. The fitted multiple regression equation for prediction of IA is given in Eqn. (14) along with the R^2 and R_{adj}^2 values.

$$E(Y_{IA}) = -181.4 + 570.4A - 297.42B + 116.15C + 208.88D + 245.26E + 484.1F - 76.74AC - 149.48AD - 35.86AE + 72.2BC - 106.06CF - 104.6EF \quad (14)$$

$$[R^2 = 0.996 \quad \text{and} \quad R_{adj}^2 = 0.994]$$

On the other hand, for prediction of probabilities of 5 categories of DC (ordinal response), 4 ordinal logistic regression equations are developed using the observed frequencies in different categories of DC. These equations are:

$$\log it I = 3.48155 + 0.63594A - 1.47767B - 1.13997C + 0.26504D - 0.14133E - 0.31945F \quad (15)$$

$$\log it II = 4.67765 + 0.63594A - 1.47767B - 1.13997C + 0.26504D - 0.14133E - 0.31945F \quad (16)$$

$$\log it III = 5.81795 + 0.63594A - 1.47767B - 1.13997C + 0.26504D - 0.14133E - 0.31945F \quad (17)$$

$$\log it IV = 6.84735 + 0.63594A - 1.47767B - 1.13997C + 0.26504D - 0.14133E - 0.31945F \quad (18)$$

Using the above logistic regression equations, the cumulative probabilities of the i^{th} category can be computed as

$$P(y_{DC} \leq i) = \frac{\exp(\log it i)}{1 + \exp(\log it i)}, \quad i = I, II, III, IV \quad (19)$$

From the cumulative probabilities, individual probability of each category can be easily computed. For the V^{th} category, the individual probability is $1 - P(y_{DC} \leq IV)$, as the cumulative probability for V^{th} category is always 1.

For the purpose of using 'Solver' tool for determination of optimal setting combination, it is now required to compute in Excel worksheet the overall desirability index for any one setting combination. Arbitrarily the existing setting combination, i.e. $A_1B_1C_3D_3E_1F_2$ is chosen here for carrying out computation in the Excel worksheet. The expected value of the quantitative response variable IA at this setting is obtained using regression equation (14). Since IA is NTB type of response variable, the desirability index for IA is computed from this expected value using equation (2). On the other hand, the expected individual probabilities of each category of the ordinal variable DC are obtained using ordinal logistic equations (15-19). Since category I is most preferred, weight 5 is assigned to category I and weights 4 to 1 are assigned to categories II to V respectively. Then using equations (9) and (10) respectively, the location score (LS) and dispersion score (DS) for the ordinal response DC are computed and then, those are converted into desirability indices using Eqns. (11-12). According to the formula, shown in Eqn. (13), geometric mean of the desirability index of IA, desirability index of LS of DC and desirability index of DS of DC is taken as the overall desirability index D.

The computed expected value of IA and the expected probabilities in different categories of DC at the existing setting combination, i.e. $A_1B_1C_3D_3E_1F_2$ are shown in Table 2. The LS and DS values for the ordinal response DC are computed as 3.6825 and 13.467 respectively. The computed desirability index for IA and desirability indices for LS and DS of DC at the existing combination are given in Table 3.

Table 2. Expected value of IA and probabilities in five categories of DC

Setting Combination	Ion Amount (IA)	Categories of DC				
	Exp value	1	2	3	4	5
$A_1B_1C_3D_3E_1F_2$	936.65	0.317	0.289	0.222	0.103	0.069

Table 3. Desirability index of IA and desirability indices for LS and DS of DC

Setting Combination	d_{IA}	d_{DC}^{LS}	d_{DC}^{DS}
$A_1B_1C_3D_3E_1F_2$	0.467	0.450	0.451

The overall desirability index, D, at the existing combination is thus calculated as

$$D = (d_{IA} \times d_{DC}^{LS} \times d_{DC}^{DS})^{1/3} = 0.456$$

The overall desirability D is essentially a function of the level values of the input variables. We now use 'Solver' optimization tool of Excel to maximize the value of D by changing the level values of input variables. Hsieh and Tong (2001) analyzed the same data using their proposed neural network (NN)-based approach and determined the optimal process condition considering no integer restrictions (IR) for the factors B-F. They also determined the optimal process condition using Taguchi methods and trade-off compromising strategy in which factors B-F are considered as integer variables. Using their NN-based approach, they derived the optimal solution as $A_1B_1C_{0.9}D_{2.5}E_3F_2$ and using Taguchi methods with trade-off compromising strategy, they obtained that the optimal combination is $A_1B_2C_1D_3E_2F_1$. In order to facilitate comparison of the optimization performance with the Hsieh and Tong's (2001) optimal solutions, solver tool is applied first without imposing integer restriction (IR) for factor B-F, and then solver tool is applied again with imposing IR on the factors B-F. By applying the proposed desirability-based approach without considering IR for factors B-F and with considering IR for factors B-F, the optimal process conditions are determined as $A_2B_1C_1D_3E_{2.83}F_1$ and $A_2B_1C_1D_3E_3F_1$ respectively. The expected value of the quantitative response IA, expected probabilities in the five categories of the ordinal response DC, and the values of the desirability indices for IA, LS of DC and DS of DC under these two optimal setting combinations of input variables are

presented in Table 4. Hsieh and Tong (2001) computed expected values of IA and probabilities in five categories of DC at the two optimal process conditions ($A_1B_1C_{0.9}D_{2.5}E_3F_2$ and $A_1B_2C_1D_3E_2F_1$) derived by them. For the purpose of easy comparison of optimization performance, the expected values of IA and probabilities in five categories of DC at the two optimal process conditions ($A_1B_1C_{0.9}D_{2.5}E_3F_2$ and $A_1B_2C_1D_3E_2F_1$) are reproduced from Hsieh and Tong (2001) in Table 4. Then, expected values of desirability indices for IA, LS of DC and DS of DC as well as overall desirability value D at these optimal conditions are also computed and presented in Table 4.

Table 4. Expected values of different measures at different optimal process conditions

Optimal solutions (Method)	IA	Exp. prob. in different categories of DC					Values of desirability indices			
	Exp. value	1	2	3	4	5	d_{IA}	d_{DC}^{LS}	d_{DC}^{DS}	D
$A_2B_1C_1D_3E_{2.83}F_1$ (Proposed approach without IR)	1000.00	0.901	0.067	0.022	0.007	0.004	1.0	0.929	0.984	0.971
$A_1B_1C_{0.9}D_{2.5}E_3F_2$ (NN-based approach without IR)	1056.8	0.88	0.08	0.03	0.01	0.00	0.513	0.917	0.977	0.772
$A_2B_1C_1D_3E_3F_1$ (Proposed approach with IR)	1011.47	0.899	0.068	0.022	0.007	0.004	0.889	0.927	0.984	0.932
$A_1B_2C_1D_3E_2F_1$ (Taguchi method & trade off with IR)	1072.6	0.76	0.09	0.09	0.05	0.01	0.406	0.783	0.921	0.664

The first two rows in Table 4 reveal the comparison of optimization performance of the proposed desirability-based approach and the Hsieh and Tong's (2001) NN-based approach, where no integer restriction is imposed for the levels of the continuous type control factors. It may be noted that under the optimal process condition derived by the proposed method, the expected value of IA becomes 1000 and the expected probability in first category (very good) of the ordinal response DC becomes 0.901. These leads to desirability values for IA, LS of DC and DS of DC as 1.0, 0.929 and 0.984 respectively, and the overall desirability D becomes as high as 0.971. On the other hand, at the optimal process condition derived by the NN-based approach (Hsieh and Tong, 2001), the expected value of IA substantially deviates from its target value leading to quite poor desirability value for IA as well as overall desirability. Thus, the results in the first two rows clearly indicate that the proposed desirability-based approach leads to better optimization performance than the NN-based approach. The results in last two rows in Table 4 reveal further that the proposed desirability-based approach (with IR for the control factors) also leads to better optimization performance than Taguchi method & trade off-based approach in terms of individual as well as overall desirability indices.

6. Conclusion

In real world, the overall quality of a product is often represented partly by the measured values of some quantitative variables and partly by the observed values of some ordinal variables. For ensuring production of high quality products, the manufacturing process of such a product needs to be optimized with respect to all these quantitative and ordinal response variables. In such a situation, usually the ordinal responses are treated as countable quantitative response variables and then, the optimal process setting is determined by applying appropriately chosen technique for simultaneous optimization of multiple quantitative response variables. But this approach, may not lead to the true optimal solution. In this article, a new approach for simultaneous optimization of quantitative and ordinal responses is presented, where ordinal response is treated as it is. In the proposed method, the relationship of the ordinal and the quantitative responses with the input controllable variables (control factors) are modeled by using ordinal logistic regression and multiple regression respectively. Then, individual desirability of the ordinal responses are computed based on the predicted probabilities for different categories of ordinal responses and individual desirability of the quantitative responses are computed based on the predicted values of the quantitative responses. Finally, an overall desirability function is defined, which is maximized to determine the optimal process settings. The results obtained from analysis of a case study data reveals that the proposed method leads to better optimization performance than the NN-based approach as well as Taguchi method-based trade off approach. After optimization of the process settings, the process needs to be monitored continuously for achieving good quality product consistently. Therefore, further research is required for developing appropriate methodology for simultaneous monitoring of quantitative and ordinal quality characteristics.

Acknowledgment

Many thanks to the referees for their advantageous comments and ideas that have significantly enhanced the substance and arrangement of this contribution.

References

- Agresti, A., 1986. Discussion of 'Testing in industrial experiments with ordered categorical data'. *Technometrics*, Vol. 28, pp. 292–294.
- Agresti, A., 2010. *Analysis of Ordinal Categorical Data*, John Wiley & Sons, Inc., New Jersey, USA.
- Ahmad, N., Kamal, S., Raza, Z.A. and Hussain, T., 2017. Multi-objective optimization in the development of oil and water repellent cellulose fabric based on response surface methodology and the desirability function. *Materials Research Express*, Vol. 4, No. 3, pp. 1-16.
- Asiabar, M.H. and Ghomi, S.M.T.F., 2006. Analysis of ordered categorical data using expected loss minimization. *Quality Engineering*, Vol. 18, pp. 117-121.
- Bashiri, M., Kamranrad, R. and Karimi, H., 2012. Response optimization in ordinal logistic regression using heuristic and meta-heuristic algorithm. *Journal of Sharif University*, Vol. 28, pp. 79-92.
- Box, G.E.P. and Jones, S., 1986. Discussion of 'Testing in industrial experiments with ordered categorical data'. *Technometrics*, Vol. 28, pp. 295–301.
- Candioti, L.V., De Zan, M.M., Cámara, M.S. and Goicoechea, H.C., 2014. Experimental design and multiple response optimization using the desirability function in analytical methods development. *Talanta*, Vol. 124, pp. 123-138.
- Castillo, E.D. and Montgomery, D.C., 1993. A nonlinear programming solution to the dual response problem. *Journal of Quality Technology*, Vol. 25, pp. 199–204.
- Castillo, E.D., Montgomery, D.C. and McCarville, D., 1996. Modified desirability functions for multiple response optimization. *Journal of Quality Technology*, Vol. 28, pp. 337–345.
- Chiang, T.L. and Su, C.T., 2003. Optimization of TQFP modelling process using neuro-fuzzy-GA approach. *European Journal of Operations Research*, Vol. 147, pp. 156–164.
- Chipman, H. and Hamada, M., 1996. Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, Vol. 35, pp. 1 – 10.
- Derringer, G. and Suich, R., 1980. Simultaneous optimization of several response variables. *Journal of Quality Technology*, Vol. 12, pp. 214–219.
- Fard, N., Xu, H. and Fang, Y., 2016. A unique solution for principal component analysis-based multi-response optimization problems. *The International Journal of Advanced Manufacturing Technology*, Vol. 82, No. 1-4, pp. 697-709.
- Ghani, S.A., Muhamad, N.A., Noorden, Z.A., Zainuddin, H. and Ahmad, A.A., 2017. Multi-response Optimization of the properties of natural ester oil with mixed antioxidants using Taguchi-based methodology. *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 24, No.3, pp. 1674-1684.
- Hamada, M. and Wu, C.F.J., 1986. Should accumulation analysis and related methods be used for industrial experiments? *Technometrics*, Vol. 28, pp. 302–306.
- Harrington, E.C. Jr., 1965. The desirability function. *Industrial Quality Control*, Vol. 21, pp. 494-498.
- Hsieh, K.L. and Tong, L.I., 2001. Optimization of multiple quality responses involving qualitative and quantitative characteristics in IC manufacturing using neural networks. *Computers in Industry*, Vol. 46, pp. 1 – 12.
- Jeng, Y.C. and Guo, S.M., 1996. Quality improvement for RC06 chip resistor. *Quality and Reliability Engineering International*, Vol. 12, pp. 439-445.
- Jeong, I-J. and Kim, K-J., 2003. Interactive desirability function approach to multi-response surface optimization. *International Journal of Reliability, Quality and Safety Engineering*, Vol. 10, No. 2, pp. 205-217.
- Khanna, R., Kumar, A., Garg, M.P., Singh, A. and Sharma, N., 2015. Multiple performance characteristics optimization for Al 7075 on electric discharge drilling by Taguchi grey relational theory. *Journal of Industrial Engineering International*, Vol. 11, No. 4, pp 459–472.
- Khuri, A.I. and Conlon, M., 1981. Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, Vol. 23, pp. 363–375.
- Kushwaha, S., Sikdar, S., Mukherjee, I. and Ray, P.K., 2013. A modified desirability function approach for mean-variance optimization of multiple responses. *International Journal of Software Science and Computational Intelligence*, Vol. 5, No. 3, pp. 7-21.
- Kim, K. and Lin, D., 2000. Simultaneous optimization of multiple responses by maximizing exponential desirability functions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 43, pp. 311-325.

- Koch, G.G., Tangen, C., Tudor, G. and Stokes, M., 1990. Strategies and issues for the analysis of ordered categorical data from multifactor studies in industry. *Technometrics*, Vol. 32, pp. 137–149.
- Kumar, P., Barua, P. and Gaindhar, J., 2000. Quality optimization (multi characteristics) through taguchi's technique and utility concept. *Quality and Reliability Engineering International*, Vol. 16, pp. 475–485.
- Liao, T., Socha, K., Oca, M.A.M., Stützle, T. and Dorigo, M., 2014. Ant colony optimization for mixed-variable optimization problems. *IEEE Transactions on Evolutionary Computation*, Vol. 18, No. 4, pp. 503 – 518.
- McCullagh, P., 1986. Discussion on 'Testing in industrial experiments with ordered categorical data'. *Technometrics*, Vol. 28, pp. 307.
- Nair, V.N., 1986. Testing in industrial experiments with ordered categorical data. *Technometrics*, Vol. 28, pp. 283–291.
- Pal, S. and Gauri, S.K., 2010. Multi-response optimization using multiple regression-based weighted signal-to-noise ratio. *Quality Engineering*, Vol. 22, pp. 336–350.
- Pan, L.K., Wang, C.C., Wei, S.L. and Sher, H.F., 2007. Optimizing multiple quality characteristics via Taguchi method-based grey analysis. *Journal of Material Processing Technology*, Vol. 182, pp. 107–116.
- Pignatiello, J. and Joseph, J., 1993. Strategies for robust multi-response quality engineering. *IIE Transactions*, Vol. 25, pp. 5–15.
- Priyadarshini, M., Pattnaik, S.K., Mishra, D., Panda, S. and Dhalmahapatra, K., 2015. Multi characteristics optimization of laser drilling process parameter using grey fuzzy reasoning method. *Materials Today: Proceedings*, Vol. 2, No. 4–5, pp. 1518-1532.
- Salmasnia, A. and Bashiri, M. 2015. A new desirability function-based method for correlated multiple response optimization. *The International Journal of Advanced Manufacturing Technology*, Vol. 76, No.5–8, pp 1047–1062.
- Taguchi, G., 1986. *Introduction to Quality Engineering*. Tokyo: Asian Productivity Organization.
- Tong, L.I. and Hsieh, K.L., 2000. A novel means of applying artificial neural networks to optimize multi-response problem. *Quality Engineering*, Vol. 13, pp. 11–18.
- Tong, L.I., Wang, C.H. and Chen, C.H., 2005. Optimisation 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.
- Vining, G.G. and Myers, R.H., 1990. Combining Taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology*, Vol. 22, pp. 38–45.
- Wu, F.C., 2005. Optimization of correlated multiple quality characteristics using desirability function. *Quality Engineering*, Vol. 17, pp. 119–126.
- Wu, F.C., 2008. Simultaneous optimization of robust design with quantitative and ordinal data. *International Journal of Industrial Engineering*, Vol. 15, No. 2, pp. 231 – 238.
- Wu, F.C. and Yeh, C.H., 2006. A comparative study on optimization methods for experiments with ordered categorical data. *Computers & Industrial Engineering*, Vol. 50, pp. 220-232.

Biographical notes

Dr. Surajit Pal is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Chennai Centre, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization.

Dr. Susanta Kumar Gauri is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Kolkata, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization.

Received July 2017

Accepted January 2018

Final acceptance in revised form February 2018