Journal of Statistical Theory and Applications
Vol. **18(3)**, September 2019, *pp.* 212–221
DOI: https://doi.org/10.2991/jsta.d.190701.001; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

Simultaneous Optimization of Multiple Responses That Involve Correlated Continuous and Ordinal Responses According to the Gaussian Copula Models

Fatemeh Jiryaie, Ahmad Khodadadi*

Department of Statistics, Shahid Beheshti University, 1983963113 G. C., Tehran, Iran

ARTICLE INFO

Article History
Received 14 Feb 2017
Accepted 14 June 2017

Keywords
Gaussian copula
Mixed outcomes
Multivariate distribution
Simultaneous optimization

ABSTRACT

This study investigates the simultaneous optimization of multiple correlated responses that involve mixed ordinal and continuous responses. The proposed approach is applicable for responses that have either an all ordinal categorical form are continuous but have different marginal distributions, or when standard multivariate distribution of responses is not applicable or does not exist. These multiple responses have rarely been the focus of studies despite their high occurrence during experiments. The copula functions have been used to construct a multivariate model for mixed responses. To resolve the computational problems of estimation under a high dimension of responses, we have estimated parameters of the model according to a pairwise likelihood estimation method. We adapted the generalized distance approach to determine settings of the factors that simultaneously optimized the mean of continuous responses and desired cumulative categories of the ordinal responses. A simulation study was used to evaluate the performance of the estimators from the pairwise likelihood approach. Finally, we presented an application of the proposed method in a real data example of a semiconductor manufacturing process.

1. INTRODUCTION

Improvements to product or process performance are important problems in pharmacology, agriculture, and industry, which is the focus of attention by manufacturers. Consequently, of particular interest is detection of optimal settings for the control factors at which the response presents certain desired characteristics. Several publications focus on optimizing a single response (Draper [1]; Hoerl [2]; Peace [3]; Fowlkes and Creveling [4]; Paul and Khuri [5]). However, numerous situations exist where multiple responses need to be simultaneously optimized. For example, in clinical trials it is important to determine the combination of drugs with maximum therapeutic effects and the lowest level of toxicity. In a semiconductor manufacturing process minimizing the defect count in the sensitive area and achieving the target amount of an ion implanted in a wafer may simultaneously need to be considered for an ion implantation process.

Most quality improvement studies research approaches for simultaneous optimization of multiple responses. The desirability function is one of the most popular methods to optimize a multi-response system (Harrington [6]; Derringer and Suich [7]; Kim and Lin [8]). This approach, turns the several responses into a single response, which results in a combined desirability. The desirability function approach is easy to apply and allows the user to make a subjective judgment on the importance of each response; however, this approach does not take into account the variance–covariance structure of the responses. Ignoring the possible correlations between the responses may be misleading and leads to incorrect optimization decisions. To overcome this difficulty, Elsayed and Chen [9], Ko *et al.* [10], Pignatiello [11], Tsui [12], and Vining [13] have proposed the use of the loss function approach to optimize multiple correlated responses. Khuri and Conlon [14] introduced an efficient optimization algorithm based on a generalized distance approach. These researchers assumed that all mean responses in the system depended on the same set of controllable variables via a polynomial regression model. The first step of their algorithm was to obtain individual optima of the estimated responses over the experimental region, after which they measured the deviation from the ideal optimum by means of a distance function expressed in terms of the estimated responses along with their variance–covariance structure. Finally, this function could be minimized to arrive at a set of suitable operating conditions.

The following regression models are used in the mentioned multiple response optimization procedure: ordinary least squares (OLS), generalized least squares (GLS), and multivariate regression (MVR), all of which are under the assumption of normal errors. OLS and GLS

*Corresponding author. Email: a_khodadadi@sbu.ac.ir

regressions consider modeling responses individually with the assumption of independent responses. OLS regression is under the homogeneity of error variances, whereas GLS is free from error variances. MVR models the responses simultaneously, by taking into consideration a correlation between the errors. Normality assumption of responses and homogeneity of error variances may be violated in situations where the responses are not normal, discrete, or exhibit heterogeneous variances. Such situations happen frequently in clinical and epidemiological studies. Mukhopadhyay and Khuri [15] have recently modified Khuri and Conlon's [14] algorithm to a generalized linear model (GLM), which can be used to handle multiple discrete responses and bypass the heterogeneity of error variances. They assume all margin distributions are the same and the joint distribution of responses are available and belong to the multivariate exponential family.

A number of different proposed methods optimize multiple continuous responses. Su and Tong [16] have used principle component analysis. However, Lai and Chang [17], Lu and Antony [18], and Tong and Su [19] used the fuzzy theorem whereas Wu and Chyu [20] suggested a mathematical programming method.

However, ordinal responses observed are observed in number of experiments due to the quality characteristic or the convenience of the measurement technique and cost-effectiveness. In the optimization of ordinal responses, Taguchi [21–23] primarily employed the accumulation analysis (AA) method. In this method the corresponding cumulative categories are defined, then the researcher determines the effects of the factor levels according to the probability distribution by the categories. Finally, the optimal control factor settings are obtained by the desired cumulative category and the important location effects are taken into consideration. Nair [24] has proposed two scoring schemes that separately detect the location and dispersion effects. Jeng and Guo [25] presented a weighted probability-scoring scheme (WPSS) to avoid the computational complexity of Nair's scoring scheme. Thy considered the location and dispersion effects. Chipman and Hamada [26] used a GLM with Bayesian estimation techniques to optimize these type of responses. Computational complexity was more than Nair's scoring scheme.

In case of mixed responses very few studies have been conducted that optimize the ordinal-continuous responses. Hsieh and Tong [27] have employed an artificial neural network technique to optimize the ordinal-continuous responses, a method that is employed with difficulty in industrial settings. Wu [28] has presented an approach based on the quality loss function of Taguchi [23] where ordinal responses can be treated as continuous responses and a weighted average quality loss is defined for the ordinal responses. This approach is easier than the approach by Hsieh and Tong [27], however there is no correlation between the responses.

In this article, we introduce an approach to simultaneously optimize mixed correlated continuous and ordinal responses. Our procedure can be easily applied when responses are all ordinal, all continuous with different types of marginal distributions, or in cases where standard multivariate distribution of responses is not applicable or does not exist. For example, when the entire marginal distribution of responses is gamma and the responses are correlated, it is difficult to determine the multivariate exponential distribution. In this approach we have used the Gaussian copula function. We extended the regression models for a bivariate mixed outcomes of De Leon and Wu [29] to the multivariate mixed discrete and continuous outcomes through pairwise fitting of models for the joint modeling of a multivariate mixed outcome based on the concept by Fieuws and Verbeke [30]. After specifying the effects of the factor levels on the mean continuous responses and probability distributions by the categories of the ordinal responses, we have adopted the generalized distance approach of Khuri and Conlon [14] to carry out the optimal control factor settings by mean of continuous responses and desired cumulative categories of ordinal responses. Copula-based dependencies, introduced in statistical literature by Skalar [31], allows one to model the dependence structure independently of marginal distributions. This approach provides an alternative and more useful representation of multivariate distribution compared to traditional approaches such as multivariate normality. Formally, copula can be defined as follows: Suppose that we have K marginal CDFs, $F_{X_1}(\cdot),\dots,F_{X_K}(\cdot)$, where $X_1,\dots,X_K$ are the random variables. Sklar's theorem states that every k-dimensional cumulative distribution $F_{X_1,\dots,X_K}(\cdot) = P(X_1 \le x_1,\dots,X_K \le x_k)$ of a random vector $(X_1,\dots,X_K)$ can be expressed by involving only the marginals $F_{X_1}(x_1),\dots,F_{X_K}(x_k)$ as $F_{X_1,\dots,X_K}(x_1,\dots,x_k) = C\left(F_{X_1}(x_1),\dots,F_{X_K}(x_k)\right)$, where C is a copula. Gaussian copulas are an important family which has been used in a variety of applications Song [32]. The P-dimensional Gaussian copula is defined as

$$C\left(F_{X_1}(x_1),\dots,F_{X_K}(x_K)\right) = \Phi_K\left(\Phi^{-1}\left\{F_{X_1}(x_1)\right\},\dots,\Phi^{-1}\left\{F_{X_K}(x_K)\right\};R\right),$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal distribution $\Phi(\cdot)$ and $\Phi_K(\cdot;R)$ is the k-dimensional standard multivariate normal distribution function. A more through overview can be found in the reference works by Joe [33] or by Nelsen [34].

We considered the real data obtained from a semiconductor manufacturing process in which the defect counted on the sensitive area (an ordinal response) and the amount of ion implanted (a continuous response) require simultaneous investigation for an ion implantation process (Hsieh and Tong [27]), as discussed in Section 6.

This paper is organized as follows: We define a multivariate model for mixed responses in Section 2. In Section 3, parameters of regression models and variance–covariance of the parameters are simultaneously estimated. The confidence region of the parameters, estimated mean of continuous responses, and estimated desired cumulative categories of ordinal responses are obtained in this section. We use all of these for the optimization algorithm. Section 4 outlines the optimization algorithm according to a generalized distance approach. In Section 5, we have conducted a simulation study to compare the performance of estimators from pairwise and full likelihood estimation. An application of the proposed optimization algorithm is described in Section 6 with a real data example. Finally, concluding remarks are presented in Section 7.

2. A MULTIVARIATE MODEL FOR MIXED RESPONSES

We consider a mixed multi-response obtained from the ith run of the experiment, $i = 1, \ldots, n$, is $Y_i = \left(Y_{i1}, \ldots, Y_{iP}, Z_{i1}, \ldots, Z_{iQ}\right)'$, where Y_{ip}s, for $p = 1, \ldots, P$ are continuous responses and Z_{iq}s, for $q = 1, \ldots, Q$ are ordinal responses with K_q levels. Underling Z_{iq} is Y^*_{iq}, a continuous latent variable, such that

$$Z_{iq} = k_q \Leftrightarrow \gamma_q^{k_q} < Y^*_{iq} \le \gamma_q^{k_{q+1}},\ k_q = 0, \ldots, K_q - 1,$$

where $\gamma_q^0 = -\infty$, $\gamma_q^{K_q} = +\infty$ and $\gamma_q = \left(\gamma_q^1, \ldots, \gamma_q^{K_q-1}\right)'$ is an unknown vector of the threshold parameters. Let cumulative distribution function (CDF) of Y_{ip} and Y^*_{iq} be $F_{Y_{ip}}\left(.; \mu_p(x_i), \theta_p\right)$ and $F_{Y^*_{iq}}\left(.; \mu^*_q(x_i), \theta^*_q\right)$ respectively, where $\mu_p(x_i) = E\left(Y_{ip}\right) = h_p\left(\delta'_p(x_i)\beta_p\right)$, $\mu^*_q(x_i) = E\left(Y^*_{iq}\right) = h_q\left(\delta^{*\prime}_q(x_i)\beta^*_q\right)$. Here, $\delta_p(x_i)$ and $\delta^*_q(x_i)$ are known vector functions of the covariate vector $x_i = (x_{i1}, \ldots, x_{im})'$, β_p and β^*_q are vectors of regression coefficient and θ_p and θ^*_q are the marginal parameters. We use the Gaussian copula function to jointly analyze these types of responses. This function has been used in a variety of applications (Song [32]) because of its flexibility and analytical tractability, and to extend the regression models for a bivariate mixed outcomes of De Leon and Wu [29] to the multivariate mixed outcomes. As it was shown in recent paper Jiryaie *et al.* [35] the joint CDF and density of Y_i can be obtained as follows:

$$P\begin{pmatrix} Y_{i1} \le y_{i1}, \ldots, Y_{iP} \le y_{iP}, \\ Z_{i1} = k_1, \ldots, Z_{iQ} = k_Q \end{pmatrix} = P\begin{pmatrix} Y_{i1} \le y_{i1}, \ldots, Y_{iP} \le y_{iP}, \\ \gamma_1^{k_1} < Y^*_{i1} \le \gamma_1^{k_{1+1}}, \ldots, \gamma_Q^{k_Q} < Y^*_{iQ} \le \gamma_Q^{k_{Q+1}}; R \end{pmatrix}$$

$$= \sum_{\varepsilon_1=0}^{1} \cdots \sum_{\varepsilon_Q=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\varepsilon_j}\, \Phi_{P^*}\begin{pmatrix} t_1, \ldots, t_P, \\ s_1^{(k_1+\varepsilon_1)}, \ldots, s_Q^{(k_Q+\varepsilon_Q)}; R \end{pmatrix}$$

and

$$f_{Y_i}\left(y_i; \theta\right) = \sum_{\varepsilon_1=0}^{1} \cdots \sum_{\varepsilon_Q=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\varepsilon_j} \frac{\partial^P}{\partial y_P \ldots \partial y_1} \Phi_{P^*}\begin{pmatrix} t_1, \ldots, t_P, \\ s_1^{(k_1+\varepsilon_1)}, \ldots, s_Q^{(k_Q+\varepsilon_Q)}; R \end{pmatrix}, \tag{1}$$

respectively, where $t_p = \Phi^{-1}\left(F_{Y_{ip}}\left(y_{ip}; \mu_p, \theta_p\right)\right)$, $p = 1, \ldots, P$, $s_q^{\left(k_q+\varepsilon_q\right)} = \Phi^{-1}\left(F_{Y^*_{iq}}\left(\gamma_q^{k_q+\varepsilon_q}; \mu^*_q, \theta^*_q\right)\right)$, $q = 1, \ldots, Q$, $P^* = P + Q$, θ is a vector parameter that contains $\beta_p, \beta^*_q, \gamma_q, \theta_p, \theta^*_q$ and R, $\Phi^{-1}(.)$ is the inverse function of the univariate standard normal CDF, while $\Phi_{P^*}(.; R)$ is the P^*-dimensional standard CDF with correlation matrix $R\left(r_{ij}\right)$ and $y_i = \left(y_{i1}, \ldots, y_{iP}, k_1, \ldots, k_Q\right)'$.

2.1. Estimation

The maximum likelihood estimator (MLE) of θ can be found by maximizing the log-likelihood function $\ell(\theta) = \sum_{i=1}^{n} log f_{Y_i}\left(y_i; \theta\right)$, with an iterative technique such as the Newton–Raphson updating scheme. Hence we should layout the density (1), let $f_{Y_p}\left(y_p\right)$ be the density of Y_p, $S_q = \Phi^{-1}\left(F_{Y^*_q}\left(Y^*_q\right)\right)$, $T_p = \Phi^{-1}\left(F_{Y_p}\left(Y_p\right)\right)$, and $R_{|T_1:T_k} = R_{T_{k+1}:S_Q|T_1:T_k}$ be the partial correlation matrix for, $T_{k+1}, \ldots, T_P, S_1, \ldots, S_Q$, after eliminating $T_1, \ldots, T_k$, for $k = 1, \ldots, P-1$. Note that

$$\Phi_{P^*}\begin{pmatrix} t_1, \ldots, t_P, \\ s_1^{(k_1+\varepsilon_1)}, \ldots, s_Q^{(k_Q+\varepsilon_Q)}; R \end{pmatrix} = \int_{-\infty}^{t_1} \phi(t)\, \Phi_{P^*-1}\begin{pmatrix} t_{2|1}, \ldots, t_{P|1}, \\ s_{1|1}^{(k_1+\varepsilon_1)}, \ldots, s_{Q|1}^{(k_Q+\varepsilon_Q)}; R_{|T_1} \end{pmatrix} dt, \tag{2}$$

where ϕ is a standard normal density. Since $\partial t_1/\partial y_1 = f_{Y_1}\left(y_1\right)/\phi\left(t_1\right)$ the differentiation of (2) with respect to y_1 is

$$\Phi_{P^*-1}\begin{pmatrix} t_{2|1}, \ldots, t_{P|1}, \\ s_{1|1}^{(k_1+\varepsilon_1)}, \ldots, s_{Q|1}^{(k_Q+\varepsilon_Q)}; R_{|T_1} \end{pmatrix} f_{Y_1}\left(y_1\right). \tag{3}$$

Next with applying (2), differentiation of function (3) with respect to y_2 is

$$\Phi_{P^*-2}\begin{pmatrix} t_{3|1:2}, \ldots, t_{P|1:2}, \\ s_{1|1:2}^{(k_1+\varepsilon_1)}, \ldots, s_{Q|1:2}^{(k_Q+\varepsilon_Q)}; R_{|T_1:T_2} \end{pmatrix} \times \frac{\phi\left(t_{2|1}\right)}{\sqrt{1 - r^2_{T_1T_2}}} \frac{f_{Y_2}\left(y_2\right)}{\phi\left(t_2\right)} \times f_{Y_1}\left(y_1\right),$$

where $\partial t_{2|1}/\partial y_2 = \phi\left(t_{2|1}\right)\partial t_2/\partial y_2$; so by repeated application of (2) we can find the densities of Y. We remark that notice the elements of $R_{|T_1:T_k}$ can be computed recursively. For example,

$$r_{T_pS_q|T_1:T_k} = \frac{r_{T_pS_q|T_1:T_{k-1}} - r_{T_kT_p|T_1:T_{k-1}} r_{T_kS_q|T_1:T_{k-1}}}{\sqrt{\left(1-r^2_{T_kT_p|T_1:T_{k-1}}\right)\left(1-r^2_{T_kS_q|T_1:T_{k-1}}\right)}},$$

for $p = k+1, \ldots, P$, $q = 1, \ldots, Q$.

In situations where the dimension of the vector variable Y is elevated, some difficulties will arise to layout the density (1). This is particularly true for high dimensional continuous vector variables due to differentiating the nested conditional normal distribution. In addition, the dimension of vector parameter θ in the joint distribution of Y will increase, which leads to computational problems to estimate the parameters. Fortunately for every dimension of Y all pair marginal density can be easily found from (1) as follows:

$$f_{Y_{p_1},Y_{p_2}}\left(y_{p_1}, y_{p_2}; \mu_{p_1}(x), \mu_{p_2}(x), \theta_{p_1}, \theta_{p_2}\right) = f_{Y_{p_1}}\left(y_{p_1}\right) f_{Y_{p_2}}\left(y_{p_2}\right) \times |R^*|^{-\frac{1}{2}} exp\left\{\frac{1}{2}\left(t_{p_1}, t_{p_2}\right)'\left(I_2 - R^{*-1}\right)\left(t_{p_1}, t_{p_2}\right)\right\},$$

$$f_{Y_p,Z_q}\left(y_p, k_q; \mu_p(x), \mu_q^*(x), \theta_p, \theta_q^*\right) = f_{Y_p}\left(y_p\right)\left[D_{r_{qp}}\left(s_q^{k_q+1}, t_p\right) - D_{r_{qp}}\left(s_q^{k_q}, t_p\right)\right],$$

$$f_{Z_{q_1},Z_{q_2}}\left(k_{q_1}, k_{q_2}; \mu^*_{q_1}(x), \mu^*_{q_2}(x), \theta^*_{q_1}, \theta^*_{q_2}\right) = \sum_{\varepsilon_{q_1}=0}^{1}\sum_{\varepsilon_{q_2}=0}^{1}(-1)^{\varepsilon_{q_1}+\varepsilon_{q_2}}\Phi_P\left(s_{q_1}^{(k_1+\varepsilon_1)}, s_{q_2}^{(k_2+\varepsilon_2)}\right).$$

Herein $q_1 \neq q_2, p_1 \neq p_2, q_1, q_2, q = 1, \ldots, Q$ and $p_1, p_2, p = 1, \ldots, P$, $R^* = \begin{pmatrix} 1 & r_{p_1p_2} \\ r_{p_2p_1} & 1 \end{pmatrix}$ and $D_r(a, b) = \Phi\left(\frac{a-rb}{\sqrt{1-r^2}}\right)$.

In order to overcome these complicated problems, we can use the pairwise likelihood estimation procedure of Fieuws and Verbeke [30]. In this approach instead of maximizing the full log-likelihood, each pairwise log-likelihood is separately maximized. Let the vector parameter of all possible pair likelihoods be $\psi = \left(\psi_1', \ldots, \psi_m'\right)'$, ψ_j represents the vector of all parameters in the jth bivariate joint model, $j = 1, \ldots, m$, $m = (P^*)(P^*-1)/2$ is the total number of possible pairs. Note the vector parameter θ of density (2) and ψ are not equivalent, elements of correlation matrix R in θ have a single counterpart in ψ, while the marginal parameters in θ have (P^*-1) counterparts in ψ. Thus

$$\hat{\theta} = A\hat{\psi}, \tag{4}$$

where A is a matrix containing the appropriate coefficients to calculate the averages and $\hat{\psi} = \left(\hat{\psi}_1', \ldots, \hat{\psi}_m'\right)'$ which $\hat{\psi}_j, j = 1, \ldots, m$ is obtained from maximizing the jth pair log-likelihood, $\ell_j\left(\psi_j\right) = \sum_{i=1}^n log f_{Y_{ij}}\left(y_{ij}; \psi_j\right)$, where $f_{Y_{ij}}\left(y_{ij}; \psi_j\right)$ is the jth pair joint density with $y_{i1} = (y_{i1}, y_{i2}), \ldots, y_m = \left(y_{i(P-1)}, y_{iP}\right)$. It can be shown that $\left(\hat{\theta} - \theta\right) \sim N(0, \Sigma)$, with $\Sigma = AJ^{-1}KJ^{-1}A'$, where J is a block-diagonal matrix with diagonal blocks J_{tt} and K is a symmetric matrix containing blocks K_{tr}, as $J_{tt} = \sum_{i=1}^n \left(\frac{\partial \ell_{it}(\psi_t)}{\partial \psi_t}\right)\left(\frac{\partial \ell_{it}(\psi_t)}{\partial \psi_t}\right)'$ and $K_{tr} = \sum_{i=1}^n \left(\frac{\partial \mathfrak{l}_{it}(\psi_t)}{\partial \psi_t}\right)\left(\frac{\partial \mathfrak{l}_{ir}(\psi_r)}{\partial \psi_r}\right)'$, $t, r = 1, \ldots, m$.

For computational convenience in the estimation step the constraints on $r_{ij} \in (-1, 1)$ can be removed with the Fisher's Z-transformation $r^*_{ij} = \frac{1}{2} log\left(\frac{1+r_{ij}}{1-r_{ij}}\right)$. So by the delta method, we have $SE\left(\hat{r}_{ij}\right) = \left(1+\hat{r}_{ij}\right)\left(1-\hat{r}_{ij}\right) SE\left(\hat{r}^*_{ij}\right)$, SE is the standard error.

Based on Wald [36] an approximate $100(1-\alpha)\%$ confidence region for $\beta = \left(\beta_1', \ldots, \beta_p', \beta_1^{\dagger'}, \ldots, \beta_Q^{\dagger'}\right)'$ with $\beta_q^{\dagger} = \left(\gamma_q^{k_q}, \beta_q^{*'}\right)'$, $q = 1, \ldots, Q$ is given by

$$C = \left\{\beta: \left(\hat{\beta} - \beta\right)'\left(var\left(\hat{\beta}\right)\right)^{-1}\left(\hat{\beta} - \beta\right)\right\} \leq \chi^2_{\alpha,\tilde{p}}\}, \tag{5}$$

here $\tilde{p}$ is the length of the vector parameter β.

In the ordinal-continuous multi-response system the goal of simultaneous optimization is determining a point, x_o, in the design region, R, at which the estimated mean responses of continuous variables, $\hat{\mu}_p(x_o) = h_p\left(\delta_p'(x_o)\hat{\beta}_p\right)$, and cumulative probabilities of desired categories $k_q = k_q^o$, $q = 1, \ldots, Q$, $\hat{p}\left(Z_q \leq k_q^o\right) = F_{Y_q^*}\left(\delta'^{\dagger}_q(x_o)\hat{\beta}_q^{\dagger}; 0, \hat{\theta}_q^*\right)$, $\delta^{\dagger}_q(x) = \left(1, -\delta'^{*}_q(x)\right)'$, are optimal. Let $\mu(x) = \left(\mu_1(x), \ldots \mu_P(x),\right.$

$p\left(Z_1 \le k_1^o\right), \ldots, p\left(Z_Q \le k_Q^o\right)\Big)'$, $\eta_p(x) = \delta_p'(x)\beta_p$ and $\eta_q^o(x) = \delta'^{\dagger}_q(x)\beta_q^{\dagger}$ a first-order Taylor expansion of $\hat{\mu}_p(x)$ and $\hat{p}\left(Z_q \le k_q^o\right)$ around $\eta_p(x)$ and $\eta_q^o(x)$, respectively are

$$\hat{\mu}_p(x) = \mu_p(x) + \frac{\partial h_p\left(\eta_p(x)\right)}{\partial \eta_p(x)}\left[\hat{\eta}_p(x) - \eta_p(x)\right]$$

and

$$\hat{p}\left(Z_q \le k_q^o\right) = p\left(Z_q \le k_q^o\right) + \frac{\partial F_{Y_q^*}\left(\eta_q^o(x); 0, \theta_q^*\right)}{\partial \eta_q^o(x)}\left[\hat{\eta}_q^o(x) - \eta_q^o(x)\right].$$

Thus an approximation of $var(\hat{\mu}(x))$ is given by

$$var(\hat{\mu}(x)) = ABvar\left(\hat{\beta}\right)B'A'. \tag{6}$$

Herein $A = diag\left(\frac{\partial\mu_1(x)}{\partial\eta_1(x)}, \ldots, \frac{\partial\mu_P(x)}{\partial\eta_P(x)}, \frac{\partial F_{Y_1^*}\left(\eta_1^o(x);0,\theta_1\right)}{\partial\eta_1^o(x)}, \ldots, \frac{\partial F_{Y_Q^*}\left(\eta_Q^o(x);0,\theta_Q\right)}{\partial\eta_Q^o(x)}\right)$, B is a block-diagonal matrix with diagonal blocks $\delta_p(x)$, $p = 1, \ldots$, P and $\delta_q^{\dagger}(x)$, $q = 1, \ldots$, Q. So the estimation of (6) is

$$\widehat{var}(\hat{\mu}(x)) = \hat{A}Bvar\left(\hat{\beta}\right)B'\hat{A}', \tag{7}$$

where $\hat{A}$ is the MLE of A.

3. THE SIMULTANEOUS OPTIMIZATION PROCEDURE

At the outset, we individually optimize each estimated mean response of continuous variables, $\hat{\mu}_p(x)$, p $= 1, \ldots$, P and cumulative probabilities of desired categories $F_{Y_q^*}\left(\delta'^{\dagger}_q(x_o)\hat{\beta}_q^{\dagger}; 0, \hat{\theta}_q^*\right)$, $q = 1, \ldots, Q$ over the experimental region R. We denote the individual optimum value of $\hat{\mu}_p(x)$, $p = 1, \ldots, P$ and $F_{Y_q^*}\left(\delta'^{\dagger}_q(x_o)\hat{\beta}_q^{\dagger}; 0, \hat{\theta}_q^*\right)$, $q = 1, \ldots, Q$ by $\hat{\kappa}_j$, $j = 1, \ldots, P^*$. If all the individual optima $\hat{\kappa}_1, \ldots, \hat{\kappa}_{P^*}$ are attained at the same setting of x in R, then the optimization problem will be solved and no further study will be required; otherwise, it is necessary to search compromise conditions favorable for all responses. To access these compromise conditions, we follow the generalized distance approach by Khuri and Conlon [14] that Mukhopadhyay and Khuri [15] adapted for the multivariate GLM situation. This approach attempts to find the conditions on x that minimizes the distance between the estimated mean responses vector $\hat{\mu}(x)$ and the individual optima vector $\hat{\kappa}(x) = \left(\hat{\kappa}_1(x), \ldots, \hat{\kappa}_{p^*}(x)\right)'$. Such a distance function is denoted by $\rho[\hat{\mu}(x), \hat{\kappa}]$. A variety of choices is possible for the distance measure ρ, Khuri and Conlon [14] have suggested the two following distance functions:

$$\rho_1[\hat{\mu}(x), \hat{\kappa}] = \left\{(\hat{\mu}(x) - \hat{\kappa})'\{\widehat{var}(\hat{\mu}(x))\}^{-1}(\hat{\mu}(x) - \hat{\kappa})\right\}^{\frac{1}{2}} \tag{8}$$

and

$$\rho_2[\hat{\mu}(x), \hat{\kappa}] = \left\{\sum_{j=1}^{m}\frac{\left(\hat{\mu}_j(x) - \hat{\kappa}_j\right)^2}{\hat{\kappa}_j^2}\right\}^{\frac{1}{2}}.$$

Since $\hat{\kappa}_1, \ldots, \hat{\kappa}_P^*$ are the individual optimum values of the random variables $\hat{\mu}_p(x)$, $p = 1, \ldots, P$ and $F_{Y_q^*}\left(\delta'^{\dagger}_q(x_o)\hat{\beta}_q^{\dagger}; 0, \hat{\theta}_q^*\right)$, $q = 1, \ldots, Q$, they are random variables themselves. We need to incorporate the variability of $\hat{\kappa}$ into ρ. For this purpose, suppose $\xi_j(\beta)$ to be the true optimum value of the jth $\left(j = 1, \ldots, P^*\right)$ mean response or cumulative probability optimized individually over R, and let $\xi(\beta) = [\xi_1(\beta), \ldots, \xi_{P^*}(\beta)]'$. Minimizing $\rho[\hat{\mu}(x), \xi(\beta)]$ is impossible because $\xi(\beta)$ is unknown. We instead minimize an upper bound of $\rho[\hat{\mu}(x), \xi(\beta)]$ to carry out the optimization step, as in Khuri and Conlon [4]. Let D_ξ be a confidence region for the true optima vector $\xi(\beta)$. By employing the $100(1-\alpha)\%$ confidence region C on β given in (5), Mukhopadhyay and Khuri [15] show that an appropriate choice of D_ξ is given by

$$P\left[\xi(\beta) \in \times_{j=1}^{P^*} D_j(C)\right] = P\left[\xi_j(\beta) \in D_j(C) \,|\, j = 1, \ldots, P^*\right] \ge P(\beta \in C) \approx 1 - \alpha, \tag{9}$$

where $\times$ denotes the Cartesian product and $D_j(C) = \left(\min_{\gamma\in C}\xi_j(\gamma), \max_{\gamma\in C}\xi_j(\gamma)\right)$ is an individual confidence interval of $\xi_j(\beta)$.

If $\xi(\beta) \in D_\xi$, then we have

$$\rho[\hat{\mu}(x), \xi(\beta)] \le \max_{\xi\in D_\xi}\rho[\hat{\mu}(x), \xi], \tag{10}$$

and hence

$$\min_{x\in R}\rho\left[\hat{\mu}(x),\xi(\beta)\right]\le\min_{x\in R}\max_{\xi\in D_\xi}\rho\left[\hat{\mu}(x),\xi\right].$$

Therefore by minimizing the right-hand side of (10) over the region R, we adopt a conservative distance approach to the optimization problem.

4. SIMULATION STUDY: FULL VERSUS PAIRWISE LIKELIHOOD ESTIMATION

In order to evaluate and compare the performance of the estimators from the pairwise and full likelihood approaches for the mixed ordinal-continuous responses, we have considered a simple 3×3 design with $x_i=(x_{i1},x_{i2})'$, $x_{ij}\in\{-1,0,1\}$, $i=1,\ldots,9$ and $j=1,2$. Distribution of the responses have the following form:

$$Y_{i1}\sim Gamma\left(\mu_1(x_i),\nu\right),\quad Y_{i2}\sim \text{normal}\left(\mu_2(x_i),\sigma^2\right),$$

where $\mu_1(x_i)=\exp(\beta_{10}+\beta_{11}x_{i1}+\beta_{12}x_{i2})$, $\mu_2(x_i)=\beta_{20}+\beta_{21}x_{i1}+\beta_{22}x_{i2}$ and the density of the $Gamma\left(\mu_1(x_i),\nu\right)$ distribution is specified as

$$f_{Y_{i1}}\left(y_{i1};\mu_1(x_i),\nu\right)=\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_1(x_i)}\right)^{\nu}y_{i1}^{\nu-1}e^{-\frac{\nu}{\mu_1(x_i)}y_{i1}},\tag{11}$$

with $E(Y_{i1})=\mu_1(x_i)$ and CDF $F_{Y_{i1}}(.;\mu_1(x_i),\nu)$,

$$Z_{i1}=\begin{cases}0 & Y_{i1}^*+\mu_1^*(x_i)\le\gamma_1\\ 1 & Y_{i1}^*+\mu_1^*(x_i)>\gamma_1\end{cases},\ Z_{i2}=\begin{cases}0 & Y_{i2}^*+\mu_2^*(x_i)\le\gamma_2^1\\ 1 & \gamma_2^1<Y_{i2}^*+\mu_2^*(x_i)\le\gamma_2^2\\ 2 & Y_{i2}^*+\mu_2^*(x_i)>\gamma_2^2\end{cases},$$

where $Y_1^*,\ Y_2^*\sim\text{normal}(0,1)$, $\mu_1^*(x_i)=\beta_{11}^*x_{i1}+\beta_{12}^*x_{i2}$, $\mu_2^*(x_i)=\beta_{21}^*x_{i1}+\beta_{22}^*x_{i2}$. Following (1) the joint density of $Y_i=(Y_{i1},Y_{i2},Z_{i1},Z_{i2})'$ can be simplified as

$$f_{Y_i}\left(y_i;\theta\right)=\frac{f_{Y_{i1}}\left(y_{i1};\mu_1(x_i),\nu\right)\phi\left(t_{2|1}\right)}{\sigma\sqrt{1-r_{T_1,T_2}^2}}\varphi\left(y_{i1},y_{i2},k_1,k_2\right),$$

herein $\varphi\left(y_{i1},y_{i2},k_1,k_2\right)$ is evaluated for $y_{i1}\ge 0$ and $y_{i2}\in R$ at $k_1=0,1$ and $k_2=0,1,2$ in Table 1 with $\Phi_2(\cdot;r^*)$ the standard bivariate normal CDF with correlation r^* and $t_2=\left(y_{i2}-\mu_2(x_i)\right)/\sigma$.

$$t_1=\Phi^{-1}\left(F\left(y_{i1};\mu_1(x_i),\nu\right)\right),\ t_{2|1}=\frac{t_2-r_{T_1T_2}t_1}{\sqrt{1-r_{T_1T_2}^2}},$$

$$s_{1|1}=\frac{\gamma_1-\mu_1^*(x_i)-r_{T_1S_1}t_1}{\sqrt{1-r_{T_1S_1}^2}},\ s_{1|1:2}=\frac{s_{1|1}-r_{T_2S_1|T_1}t_{2|1}}{\sqrt{1-r_{T_2S_1|T_1}^2}},$$

$$s_{2|1}^{(j)}=\frac{\gamma_2^{(j)}-\mu_2^*(x_i)-r_{T_1S_2}t_1}{\sqrt{1-r_{T_1S_2}^2}},\ s_{2|1:2}^{(j)}=\frac{s_{2|1}^{(j)}-r_{T_2S_2|T_1}t_{2|1}}{\sqrt{1-r_{T_2S_2|T_1}^2}},$$

Table 1 | $\varphi\left(y_{i1},y_{i2},k_1,k_2\right)$ evaluated at $k_1=0,1$, and $k_2=0,1,2$.

	$k_1=0$	$k_1=1$
$k_2=0$	$\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(1)};r^*\right)$	$\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(2)};r^*\right)-\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(1)};r^*\right)$
$k_2=1$	$\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(2)};r^*\right)-\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(1)};r^*\right)$	$\Phi\left(s_{2\|1:2}^{(2)}\right)-\Phi\left(s_{2\|1:2}^{(1)}\right)-\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(2)};r^*\right)+\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(1)};r^*\right)$
$k_2=2$	$\Phi\left(s_{1\|1:2}\right)-\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(2)};r^*\right)$	$1-\Phi\left(s_{2\|1:2}^{(2)}\right)-\Phi\left(s_{1\|1:2}\right)+\Phi_2\left(s_{1\|1:2},s_{2\|1:2}^{(2)};r^*\right)$

Table 2 Result of simulation study for pairwise and full likelihood estimation.

Parameter	Truth	$RB_p \times 100$	$RB_f \times 100$	Ave SE_p	SD_p	RE_p	MSE_p	MSE_f
β_{10}	1	−0.8577	−25.0141	0.1035	0.1107	0.9354	0.0123	0.0798
β_{11}	−3	−0.1182	−0.2141	0.1260	0.1559	0.8080	0.0243	0.0254
β_{12}	−2	−0.2783	−1.1402	0.1259	0.1374	0.9164	0.0189	0.0217
β_{20}	−1	3.6322	16.4251	0.1867	0.1641	1.1380	0.0282	0.0608
β_{21}	1	0.7144	−0.1973	0.2410	0.2026	1.1893	0.0411	0.0451
β_{22}	−1	−1.4872	−2.2351	0.2396	0.2007	1.1939	0.0405	0.0454
γ_1	1	11.4930	−29.3023	0.3426	0.3477	0.9854	0.1272	0.0890
β^*_{11}	−1	10.1537	−10.6593	0.3854	0.3956	0.9741	0.1513	0.0567
β^*_{12}	2	10.6494	−14.6922	0.5231	0.4963	1.0539	0.3493	0.0919
γ^1_1	−0.25	1.6823	116.7175	0.2153	0.1948	1.1054	0.0380	0.0880
γ^2_1	1	6.0361	−2.0552	0.2547	0.2356	1.0808	0.0592	0.0396
β^*_{21}	1	6.4798	−15.4204	0.2652	0.2426	1.0930	0.0631	0.0521
β^*_{22}	−2	5.1658	−14.7951	0.3553	0.3184	1.1161	0.1120	0.0920
σ	1	−4.3821	29.3046	0.0753	0.1196	0.6298	0.0162	0.0890
ν	0.4	3.5435	−20.1926	0.0534	0.0521	1.0258	0.0029	0.0090
$r_{T_1T_2}$	0.5	−1.5025	18.1566	0.0918	0.0851	1.0793	0.0073	0.0098
$r_{T_1S_1}$	0.5	4.4983	−40.5024	0.1917	0.1621	1.1827	0.0268	0.0424
$r_{T_1S_2}$	0.5	2.7490	−40.6943	0.1427	0.1278	1.1160	0.0165	0.0428
$r_{T_2S_1}$	0.5	5.1281	−31.6022	0.2037	0.1746	1.1671	0.0311	0.0331
$r_{T_2S_2}$	0.5	3.3858	−33.5907	0.1484	0.1312	1.1306	0.0175	0.0340
$r_{S_1S_2}$	0.5	6.8221	18.7066	0.3100	0.2837	1.0927	0.0817	0.0099

True values are given, the relative bias (RB = Bias ÷ Parameter) under the pairwise $\left(RB_p\right)$ and full likelihood $\left(RB_f\right)$, average estimate of standard errors $\left(\text{Ave } SE_p\right)$, empirical standard division $\left(SD_p\right)$ and relative efficiency $\left(RE_p = \text{Ave } SE_p \div SD_p\right)$ under the pairwise approach are obtained. Mean squared error under pairwise $\left(MSE_p\right)$ and full likelihood $\left(MSE_f\right)$ are listed.

$$r_{T_2S_j|T_1} = \frac{r_{T_2S_j} - r_{T_1T_2} r_{T_1S_j}}{\sqrt{\left(1 - r^2_{T_1T_2}\right)\left(1 - r^2_{T_1S_j}\right)}}, \quad r^* = \frac{r_{S_1S_2|T_1} - r_{T_2S_1|T_1} r_{T_2S_2|T_1}}{\sqrt{\left(1 - r^2_{T_2S_1|T_1}\right)\left(1 - r^2_{T_2S_2|T_1}\right)}},$$

for $j = 1, 2$. A total of $R = 1000$ repeated samples of size $n_i = 10$, $i = 1, \ldots, 9$ for each run of experiments were generated with the true values of parameters given in Table 2. Estimation of parameters and calculation of the likelihood score functions at estimated parameters were implemented in R using the "optim" and "fdHess" functions, respectively. The results of this simulation study are reported in Table 2. Relative biases and mean square errors of the pairwise and full likelihood suggest that the pairwise likelihood estimation obtains suitable point estimates with small mean square errors. Furthermore relative efficiencies that are generally close to 1 in the pairwise method show that pairwise likelihood estimations have standard errors which reflect these estimates true sampling variability.

5. ILLUSTRATIVE EXAMPLE

In dealing with simultaneous optimization of mixed ordered categorical and continuous responses, a case study of an ion implantation process from a Taiwanese integrated circuit (IC) fabrication manufacturer was conducted by Hsieh and Tong [27] based on artificial neural networks. This example contained two quality responses: i) the amount of ion implanted in a wafer, continuous response denoted by Y, and ii) the defect situation of a sensitive area in the wafer, an ordered response denoted by Z, which included five ordered categories: very good, good, not good not bad, bad, and very bad. The responses was listed in a progressively worse order; we denoted these categories as 1–5. Each wafer had 36 sensitive areas that were tested independently. There were six control factors denoted by $X_1, \ldots, X_6$. Between these, X_1 was discrete whereas the others were continuous. Table 3 lists the control factors with their levels and coded levels of $X_1, \ldots, X_6$ denoted by $x_1, \ldots, x_6$. Therefore the region of the experiment transformed to the $R = \left\{x = (x_1, \ldots, x_6)' \,|x_1 = 0, 1; 1 \le x_l \le 3, l = 2, ..6\right\}$. The two mixed responses data in a L_{18} orthogonal array are given in Table 4. In this table m_{ik}, $i = 1, \ldots, 18$, $k = 1, ..5$ are the number of 36 sensitive areas in the ith wafer, which was tested in the ith run (ith level of x) of the experiment which fell into the kth category.

The continuous response is a nominal-the-best (NTB) with a target value of 1000 (after the data was transformed). First, for the continuous response Y we fitted the normal GLM regression with link $\mu(x_i) = \delta'(x_i)\beta$ and gamma GLM regression with link $\mu(x_i) = \exp(\delta'(x_i)\beta)$

Table 3 Control factors and coded factors with their levels.

Level	X_1	X_2	X_3	X_4	X_5	X_6	x_1	x_2	x_3	x_4	x_5	x_6
Level 1	Type 1	6	50	5	4	25	1	1	1	1	1	1
Level 2	Type 2	12	100	10	8	50	0	2	2	2	2	2
Level 3		18	150	15	12	75		3	3	3	3	3

Table 4 Experimental data.

i	y_{i1}	m_{i1}	m_{i2}	m_{i3}	m_{i4}	m_{i5}
1	741.4	33	3	0	0	0
2	972.1	24	5	6	1	0
3	796.1	6	2	20	8	0
4	797.8	0	28	4	4	0
5	796.6	2	2	4	12	16
6	802.1	4	0	20	4	8
7	908.2	0	2	6	14	14
8	645.7	10	2	8	4	12
9	650.3	0	0	0	24	12
10	1072.5	34	0	2	0	0
11	1316.1	30	2	4	0	0
12	890.5	10	10	12	0	4
13	886.6	14	8	10	4	0
14	826.5	8	16	12	0	0
15	800.1	0	8	6	4	18
16	816.1	18	12	6	0	0
17	824.2	10	6	0	4	16
18	735.6	0	4	2	6	24

where the density of the $Gamma\,(\mu_i, \nu)$ distribution was specified as (11), $\delta\,(x_i) = (1, x_{i1}, \ldots, x_{i6})'$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_6)'$. Akaike information criteria (AIC) were 230.37 and 225.1, respectively. Gamma distribution appeared to have a better fit to this response compared to normal distribution. We assumed that the dispersion parameter ν was known and did not need to be estimated in the joint regression model. In our example we used its MLE in the marginal gamma GLM, which was 35.9348.

Z_i, $i = 1, \ldots, n$, $n = 18$ has the ordinal form with K = 5 categories. Therefore we have used the cumulative probabilities. The kth cumulative probability in the ith run of the experiment is

$$p\,(Z_i \leq k) = F_{Y^*}\left(\gamma_k - \delta^*\,(x_i)'\,\beta^*\right) = \frac{\exp\left(\gamma_k - \delta^*\,(x_i)'\,\beta^*\right)}{1 + \exp\left(\gamma_k - \delta^*\,(x_i)'\,\beta^*\right)},$$

where $k = 0, \ldots, (K-1)$, $\delta^* = (x_{i1}, \ldots, x_{i6})'$, $\beta^* = \left(\beta_1^*, \ldots, \beta_6^*\right)'$, and $\gamma = (\gamma_1, \ldots \gamma_4)'$.

$Y_i = (Y_i, Z_i)'$ given covariate $x_i = (x_{i1}, \ldots, x_{i6})'$, $i = 1, \ldots, 18$ are independent, $Y_i \in R^+$ and $Z_i \in \{0, \ldots, 4\}$, using (1), we can write the joint density of Y_i and Z_i as follows:

$$f_{Y_i, Z_i}\left(y_i, k; \theta\right) = f_{Y_i}\left(y_i; \mu\,(x_i)\right)\left[D_r\left(s^{k+1}, t\right) - D_r\left(s^k, t\right)\right],$$

where $t = \Phi^{-1}\left(F_{Y_i}\left(y_i; \mu\,(x_i), \nu\right)\right)$, $s^k = \Phi^{-1}\left(F_{Y_i^*}\left(\gamma_{k_i} - \mu^*\,(x_i)\right)\right)$, $D_r\,(a, b) = \Phi\left(\frac{a - rb}{\sqrt{1-r^2}}\right)$ and $\theta = (\beta', \beta'^*, \gamma', r)'$.

Therefore log-likelihood for the presented data set in Table 4 is

$$\ell\,(\theta) = \sum_{i=1}^{n} \log f_{Y_i}\left(y_i; \mu\,(x_i)\right) + \sum_{i=1}^{n}\sum_{k=0}^{K-1} m_{i(k+1)} \log\left[D_r\left(s^{k+1}, t\right) - D_r\left(s^k, t\right)\right]. \tag{12}$$

Following this log-likelihood, we estimated the parameters and calculated the likelihood's score functions at the estimated parameters in R software using the "optim" and "fdHess" functions, respectively. Table 5 lists the estimated parameters, their standard errors and p-values. In this table $\hat{r} = 0.0092$ with a p-value = 0.4672 shows we can accept that the amount of ion implanted and the defect situation of a sensitive area in a wafer are independent. In this example we aim to reach a point in the design region that simultaneously minimize $|\hat{\mu}\,(x) - 1000|$ and maximize $\hat{\pi}_1\,(x)$ as cumulative probabilities of the desired category, herein $\hat{\pi}_k\,(x) = \hat{p}\,(Z = k - 1)$, $k = 1, \ldots, 5$. Table 6 shows $\hat{\kappa}_1$ and $\hat{\kappa}_2$, the individual optimum of $\hat{\mu}\,(x)$ and $\hat{\pi}_1\,(x)$ over R, do not have the same location in the design region. So its needed to continue, we know the components of $\hat{\kappa}\,(x)$ are random variables, using these variablity in $\hat{\kappa}$ a confidence region such as (12) can be constructed, before using the distance metric. For this purpose 5000 values were randomly selected from C, the confidence region of $\left(\beta', \beta^{*'}\right)'$, and individual optimum values, $(\mu\,((x)\,, \pi_1\,(x))'$, of each point over the experimental region R were computed. We denote these optimum by $\left(\xi_1, \xi_2\right)'$ and

Table 5 Estimates and standard errors.

Parameter	Estimate	Std.Error	$z = \frac{\hat{\theta}}{se\left(\hat{\theta}\right)}$	p-value
$\hat{\beta}_0$	7.1070	0.2268	31.3394	0.0000
$\hat{\beta}_1$	−0.1316	0.0789	−1.6682	0.0476
$\hat{\beta}_2$	−0.1143	0.0481	−2.3778	0.0087
$\hat{\beta}_3$	−0.0557	0.0488	−1.1416	0.1268
$\hat{\beta}_4$	−0.0115	0.0474	−0.2426	0.4042
$\hat{\beta}_5$	−0.0281	0.0487	−0.5772	0.2819
$\hat{\beta}_6$	0.0566	0.0488	1.1610	0.1228
$\hat{\beta}_1^*$	0.7580	0.1561	4.8551	0.0000
$\hat{\beta}_2^*$	1.4816	0.1073	13.8052	0.0000
$\hat{\beta}_3^*$	1.1839	0.1009	11.7374	0.0000
$\hat{\beta}_4^*$	−0.2319	0.0939	−2.4699	0.0068
$\hat{\beta}_5^*$	0.1324	0.0970	1.3653	0.0861
$\hat{\beta}_6^*$	0.3071	0.0942	3.2607	0.0006
$\hat{\gamma}_1$	4.9282	0.4683	10.5234	0.0000
$\hat{\gamma}_2$	5.9847	0.4880	12.2635	0.0000
$\hat{\gamma}_3$	7.2068	0.5166	13.9513	0.0000
$\hat{\gamma}_4$	8.2651	0.5351	15.4455	0.0000
$\hat{r}$	0.0092	0.1125	0.0822	0.4672

Table 6 The individual optima and confidence region D_ξ.

	Gamma Response	Ordinal Response
Location	(1, 6.002, 63.848, 6.08, 4.401, 71.88)	(0, 6, 50, 15, 4, 25)
Optimum mean response	1000	0.925
Confidence region	(822.078, 1000)	(0.849, 0.965)

Table 7 Simultaneous optima (Example 1).

	Taghuchi	Neural Network	Gaussian Copula
Location	(1, 12, 50, 15, 8, 25)	(1, 6.06, 46.32, 12.19, 11.63, 52.06)	(0,7.26, 54.9, 12.55, 11.91, 25.42)
Ordinal response	(0.54, 0.23, 0.15, 0.05, 0.03)	(0.75, 0.14, 0.07, 0.02, 0.01)	(0.85,0.094, 0.04, 0.01, 0.01)
Gamma response	778.113	912.682	946.277
Maxρ_2	0.496	0.245	0.194

we find the minimum and maximum 5000 values of $\left(\xi_1, \xi_2\right)'$ to get the approximate lower and upper bound of $D_i\,(C)$, i = 1, 2, where in Table 6 $D_1\,(C) = (822.0784, 1000)$ and $D_2\,(C) = (0.8486, 0.9648)$ are confidence intervals with a 95% convergence probability for μ and π_1, respectively.

In this example we used the distance measure ρ_2 due to two responses that had an equal importance in the ion implantation process by Hsieh and Tong [27] and were independent. For each of the 5000 values of x randomly selected from the region R, we computed the maximum distance function $\rho_2\,[\hat{\mu}\,(x)\,, \xi]$ with respect to $\xi \in D_\xi$, where $D_\xi = \times_{j=1}^{2} D_j\,(C)$,and $\hat{\mu}\,(x) = (\hat{\mu}\,(x)\,, \hat{\pi}_1\,(x))'$. Next a minimum of 5000 values of this maximum distance were obtained. This distance, the corresponding simultaneous maximum of $\hat{\mu}\,(x)$ and $\hat{\pi}_1\,(x)$ and their locations are given in Table 7.

For the purpose of compression we considered two points of the design region that Hsieh and Tong [27] introduced for simultaneous optimization of these responses with the Taguchi and Artificial Neural Network methods. By using these locations, estimated parameters in Table 5 and confidence region in Table 6, $\hat{\mu}$, $(\hat{\pi}_1, \ldots, \hat{\pi}_5)'$ and maximum of $\rho_2\,[\hat{\mu}\,(x)\,, \xi]$ with respect to this confidence region were computed (Table 7). These results showed that the location founded by the Gaussian copula had better optimum values for these responses.

6. CONCLUSION

In the simultaneous optimization problem, due to the inherent nature of the data and convenience of measurements, it is not feasible to report all of the responses as continuous variables with normal distribution. One of the most popular means is to represent the data in the ordinal categorical form. Thus the outputs may involve mixed continuous and ordinal variables. The innovative use of the copula function permits a model of various types of correlated responses, such as mixed continuous and ordinal responses, those with all ordinal categorical forms, continuous responses that have different marginal distributions, or where standard multivariate distribution of the responses is not applicable or does not exist. This paper used the pairwise likelihood estimation method for a high dimension of responses, and alleviated the computational demands of estimation. The results of the simulation study showed the usefulness of this method. Adopting the generalized distance approach would allow us to simultaneously optimize such responses by considering the dependency between them. An example demonstrated the effectiveness of the proposed method. The published methods could not be directly applied to simultaneous optimization of such correlated responses.

REFERENCES

1. N.R. Draper, Technometrics. 5 (1963), 469–479.
2. A.E. Hoerl, Chem. Eng. Prog. Symposium Ser. 60 (1964), 67–77.
3. G.S. Peace, Taguchi Methods: A Hands-on Approach, Addison-Wesley, Reading, 1993.
4. W.Y. Fowlkes, C.M. Creveling, Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development, Addison-Wesley, Reading, 1995.
5. S. Paul, A.I. Khuri, Commun. Stat. Theory Methods. 29 (2000), 2181–2200.
6. E.C. Harrington, Ind. Qual. Control. 21 (1965), 494–498.
7. G. Derringer, R. Suich, J. Qual. Technol. 12 (1980), 214–219.
8. K.J. Kim, D.K.J. Lin, Appl. Stat. J. R. Stat. Soc. Ser. C. 49 (2000), 311–325.
9. E.A. Elsayed, A. Chen, Int. J. Prod. Res. 31 (1993), 1117–1132.
10. Y.H. Ko, K.J. Kim, C.H. Jun, J. Qual. Technol. 37 (2005), 50–59.
11. J. Pignatiello, IIE Trans. 25 (1993), 5–15.
12. K.L. Tsui, Int. J. Prod. Res. 37 (1999), 433–445.
13. G.G. Vining, J. Qual. Technol. 30 (1998), 309–313.
14. A.I. Khuri, M. Conlon, Technometrics. 23 (1981), 363–375.
15. S. Mukhopadhyay, A.I. Khuri, Comput. Stat. Data Anal. 52 (2008), 4625–4634.
16. C.T. Su, L.I. Tong, Total Qual. Manag. 8 (1997), 409–416.
17. Y.J. Lai, S.I. Chang, Fuzzy Sets Syst. 63 (1994), 117–129.
18. D. Lu, J. Antony, Int. J. Prod. Res. 40 (2002), 1613–1625.
19. L.I. Tong, C.T. Su, Qual. Reliab. Eng. Int. 13 (1997), 25–34.
20. F.C. Wu, C.C. Chyu, Int. J. Prod. Res. 42 (2004), 337–354.
21. G. Taguchi, Taguchi Methods: Case Studies from the U.S. and Europe, American Supplier Institute, Michigan, 1991.
22. G. Taguchi, Taguchi Methods: Research and Development, American Supplier Institute, Michigan, 1991.
23. G. Taguchi, Taguchi Methods: Signal-to-Noise Ratio for Quality Evaluation, American Supplier Institute, Michigan, 1991.
24. V.N. Nair, Technometrics. 28 (1986), 283–291.
25. Y.C. Jeng, S.M. Guo, Qual. Reliab. Eng. Int. 12 (1996), 439–445.
26. H. Chipman, M. Hamada, Technometrics. 38 (1996), 1–10.
27. K. L. Hsieh, L.I. Tong, Comput. Ind. 46 (2001), 1–12.
28. F.C. Wu, Int. J. Ind. Eng. 15 (2008), 231–238.
29. A.R. De Leon, B. Wu, Stat. Med. 30 (2011), 175–185.
30. S. Fieuws, G. Verbeke, Biometrics. 62 (2006), 424–431.
31. A. Sklar, Fonctions de Répartition à n Dimensions et Leurs Marges, Publications de l'Institut Statistique de l'Université de Paris, Paris. 8 (1959), 229–231.
32. P.X.K. Song, Correlated Data Analysis: Modeling, Analytics, and Applications, Springer, New York, 2007.
33. H. Joe, Multivariate Models and Dependence Concepts, Chapman and Hall, London, 1997.
34. R.B. Nelsen, An Introduction to Copulas, Springer, New York, 1999.
35. F. Jiryaie, N. Witanage, B. Wu, A.R. De Leon, J. Stat. Comput. Simul. 86 (2015), 1643–1659.
36. A. Wald, Trans. Am. Math. Soc. 54 (1943), 426–482.