CrossMark

Application of kernel principal component analysis to multi-characteristic parameter design problems

Woojin Soh[1] · Heeyoung Kim[1] · Bong-Jin Yum[1]

Abstract The Taguchi method for robust parameter design traditionally deals with single characteristic parameter design problems. Extending the Taguchi method to the case of multi-characteristic parameter design (MCPD) problems requires an overall evaluation of multiple characteristics, for which the principal component analysis (PCA) has been frequently used. However, since the PCA is based on a linear transformation, it may not be effectively used for the data with complicated nonlinear structures. This paper develops a kernel PCA-based method that allows capturing nonlinear relationships among multiple characteristics in constructing a single aggregate performance measure. Applications of the proposed method to simulated and real experimental data show the advantages of the kernel PCA over the original PCA for solving MCPD problems.

Keywords Kernel principal component analysis · Multiple performance characteristics · Robust parameter design · SN ratio · Taguchi method

1 Introduction

The objective of the Taguchi parameter design (Taguchi and Yokoyama 1993) is to determine optimal settings of design parameters such that the performance characteristics become robust to uncontrollable noise variables. It utilizes orthogonal arrays as experimental designs to study a large number of variables with a relatively small number of experimental runs. Once experimental data are collected, a performance measure called the signal-to-noise (SN) ratio

✉ Heeyoung Kim
heeyoungkim@kaist.ac.kr

Woojin Soh
wjsoh@kaist.ac.kr

Bong-Jin Yum
bjyum@kaist.ac.kr

[1] Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

is calculated for each run and used to determine the optimal settings of design parameters. Taguchi method is traditionally used to deal with single characteristic parameter design (SCPD) problems. Even though problems with multiple characteristics are more common in practice, they are much more difficult to solve compared to SCPD problems. In fact, a higher SN ratio for one characteristic may correspond to a lower SN ratio for another in determining optimal settings of design parameters. Consequently, an overall evaluation of multiple SN ratios is required for solving multi-characteristic parameter design (MCPD) problems.

Principal component analysis (PCA) has been widely used for solving MCPD problems. PCA is a dimensionality reduction technique that linearly transforms a number of possibly correlated variables into a small number of uncorrelated variables called principal components (Jolliffe 2002). For solving MCPD problems, Su and Tong (1997) applied PCA to multiple SN ratios and then selected principal component scores which correspond to the eigenvalues greater than 1 as aggregate performance measures. However, if more than one principal component score is selected, the problem still involves multiple performance measures, not a single aggregated one. Moreover, the ignored principal component scores which correspond to the eigenvalues less than 1 might still contain useful information. To overcome these problems, Jean and Wang (2006) used all principal component scores to construct a single aggregate performance measure by taking the logarithm of the sum of all principal component scores. Liao (2006) proposed the weighted sum of principal component scores as a single aggregate performance measure. The weight of each principal component score is chosen as the ratio of the variance of each principal component to the total variance of all principal components. Datta et al. (2009) proposed a single aggregate performance measure by taking the geometric mean of principal component scores. Sibalija and Majstorovic (2009) performed the grey relational analysis (GRA) (Wang et al. 1996), another popular dimensionality reduction method, on all principal component scores to construct a single aggregate performance measure.

The above PCA-based methods have shown good performance in solving MCPD problems by converting them to SCPD ones. However, if the data have complicated structures that cannot be well represented in a linear subspace, the original PCA may not work well for MCPD problems. To address this problem, a kernel PCA-based method is proposed in this paper. The kernel PCA (Schölkopf et al. 1998), a generalization of the original PCA, allows nonlinear feature extraction using kernel methods. By employing the kernel PCA, an effective single aggregate performance measure can be constructed even if multiple performance characteristics are nonlinearly related with each other.

Even though this paper focuses on PCA-based approaches, it is worth mentioning other approaches for MCPD problems. The GRA was first employed in Lin and Tarng (1998) and has been applied for solving MCPD problems in various fields by Lin and Lin (2002), Tarng et al. (2002), Lin (2004), Tzeng et al. (2009), Jung and Kwon (2010) and Yang et al. (2014). Artificial intelligence techniques have been also used, including artificial neural network (ANN) (Sukthomya and Tannock 2005; Tsao and Hocheng 2008), genetic algorithm (GA) (Forouraghi 2000; Jeyapaul et al. 2006; Yildiz et al. 2007), and fuzzy theory (Tong and Su 1997; Lin et al. 2000; Sharma et al. 2011). Some dimensionality reduction approaches have been combined with another as a hybrid approach. They include a PCA-GRA approach by Sibalija and Majstorovic (2009), ANN-Fuzzy approach by Antony et al. (2006), ANN-PCA approach by Hsu (2001), GRA-Fuzzy approach by Lin and Lin (2005), ANN-GA approach by Huang and Tang (2006), and PCA-GRA-ANN-GA approach by Sibalija and Majstorovic (2012).

The remainder of this paper is organized as follows. Section 2 introduces the PCA and kernel PCA. A brief account of the PCA is given in Sect. 2.1. A basic concept of the kernel

function and the kernel PCA algorithm are introduced in Sects. 2.2 and 2.3, respectively. Then, the kernel PCA-based method for MCPD problems is proposed in Sect. 3. The proposed method is applied to simulated and real experimental data in Sects. 4 and 5, respectively. The paper concludes with a discussion in Sect. 6.

2 PCA and kernel PCA

2.1 PCA

Using orthogonal linear projection, PCA transforms a set of observations of possibly correlated variables into a set of the values of uncorrelated variables called principal components. The first principal component is required to have the largest possible variance, and the second principal component has the second largest variance not represented by the first principal component, and so on. The values of these new variables are called principal component scores which are actually the projections of the observations onto the principal component space. By representing data with only the first few principal components, the dimensionality of the data can be reduced.

PCA is performed as follows. Let $\mathbb{R}^p$ be the space of p-dimensional real vectors. Suppose that $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})' \in \mathbb{R}^p$ is centered as

$$\tilde{\mathbf{x}}_i = (x_{i1} - \bar{x}_1, x_{i2} - \bar{x}_2, \ldots, x_{ip} - \bar{x}_p)', i = 1, \ldots, n$$

where $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, j = 1, \ldots, p$. A sample variance-covariance matrix of $\tilde{\mathbf{x}}_i$ is defined as

$$\hat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{x}}} = \frac{1}{n-1} \sum_{i=1}^{n} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i'.$$

Then, the eigenvalue problem for $\hat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{x}}}$ is constructed as follows.

$$\hat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{x}}} \mathbf{v}_j = \lambda_j \mathbf{v}_j, j = 1, \ldots, p \tag{1}$$

where λ_j and $\mathbf{v}_j = (v_{1j}, v_{2j}, \ldots, v_{pj})'$ are the jth largest eigenvalue and the corresponding eigenvector of $\hat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{x}}}$, respectively. $\mathbf{v}_j$ is also called the jth principal component. Then, $\tilde{\mathbf{x}}_i$ is transformed by orthogonal linear projection as follows.

$$t_{ij} = \mathbf{v}_j' \tilde{\mathbf{x}}_i = v_{1j} \tilde{x}_{i1} + v_{2j} \tilde{x}_{i2} + \cdots + v_{pj} \tilde{x}_{ip}, j = 1, \ldots, p$$

where t_{ij} is a linearly projected value on $\mathbf{v}_j$ and is called a principal component score.

When the data are geometrically distributed as an ellipsoidal form (e.g., normal distribution), it can be easily represented by linear principal components. In practice, however, the data may have a complicated structure that cannot be explained well enough by conventional linear principal components. An example of this situation is illustrated in Fig. 1a in which the first principal component of the data is indicated as a solid line. It appears that the first principal component does not adequately explain the major structure of the data. In other words, when the data have complicated structures, the original PCA, which only allows linear extraction, may result in an unreasonable representation of the data. The kernel PCA is a nonlinear generalization of the original PCA. In the kernel PCA, the data in the original space are mapped into a higher dimensional feature space where the data can be linearly modeled (see Fig. 1b). Then, the original PCA can be performed in the feature space via the
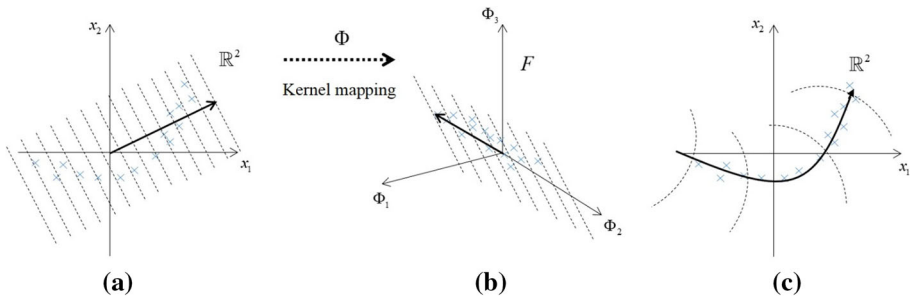
Fig. 1 Basic idea of kernel PCA

so called kernel trick (see Sect. 2.2), and the linearly extracted principal components in the feature space can better explain the nonlinear structure of the data in the original space (see Fig. 1c).

2.2 Kernel functions and kernel tricks

If the data in the original input space are transformed into potentially much higher dimensional feature space through a nonlinear mapping, then the nonlinear relations of the data in the original space may be discovered using linear learning algorithms in the feature space. Define such a nonlinear mapping Φ as follows.

$$\Phi : \mathbb{R}^p \to F$$
$$\mathbf{x} \mapsto \Phi(\mathbf{x})$$

where F is the feature space, the dimension of which is higher than that of the original space, $\mathbf{x}$ is the p-dimensional input data, and $\Phi(\mathbf{x})$ is the mapped data by a nonlinear mapping function $\Phi(\cdot)$.

In general, the feature space has a very high or even infinite dimension, and therefore, it is cumbersome to construct the mapping function $\Phi(\cdot)$ and evaluate the mapped data. Fortunately, many linear algorithms including the PCA can be reformulated in such a way that the inner product arises naturally (Bishop 2006). The inner product in the feature space can be calculated directly as a function of the original input. Consider the following mapping.

$$\Phi : (x_1, x_2)' \to (x_1^2, \sqrt{2}x_1x_2, x_2^2)'$$

Then, the inner product in the feature space can be reformulated in terms of an algebraic expression in the original space as follows (Müller et al. 2001).

$$\begin{aligned}
\Phi(\mathbf{x})'\Phi(\mathbf{y}) &= \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right)\left(y_1^2, \sqrt{2}y_1y_2, y_2^2\right)' \\
&= \left((x_1, x_2)(y_1, y_2)'\right)^2 \\
&= \left(\mathbf{x}'\mathbf{y}\right)^2 \\
&= k(\mathbf{x}, \mathbf{y})
\end{aligned} \tag{2}$$

where $k(\mathbf{x}, \mathbf{y})$ is called a kernel function. The inner product of mapped data is replaced by the kernel function in the original space, and therefore, the inner product in the feature space can be calculated in the original space without performing the nonlinear mapping $\Phi(\cdot)$. This procedure is called the kernel trick (Schölkopf and Smola 2002).

Table 1 Common kernel functions

Name	Kernel functions	Parameters
Linear kernel	$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{y}$	
Homogeneous polynomial kernel	$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}'\mathbf{y})^d$	$d \geq 2$
Non-homogeneous polynomial kernel	$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}'\mathbf{y} + c)^d$	$d \geq 2, c > 0$
Gaussian kernel	$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{y}||^2}{2\sigma^2}\right)$	$\sigma > 0$

There are various kernel functions, and the performance of kernel-based algorithms depends on the type of kernel functions. Commonly used kernel functions are summarized in Table 1. The Mercer's theorem guarantees that there exists a mapping $\Phi(\cdot)$ which satisfies $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})'\Phi(\mathbf{y})$ for such kernel functions (Müller et al. 2001).

2.3 Kernel PCA

The kernel PCA algorithm proceeds as follows Schölkopf et al. (1998). Suppose that $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$, are mapped into a feature space as $\Phi(\mathbf{x}_i)$, $i = 1, \ldots, n$. Assume further that $\Phi(\mathbf{x}_i)$'s are centered. For notational convenience, the same notation $\Phi(\mathbf{x}_i)$, $i = 1, \ldots, n$, will be used to denote $\Phi(\mathbf{x}_i)$'s after centering. Then, the following holds.

$$\sum_{i=1}^{n} \Phi(\mathbf{x}_i) = \mathbf{0} \tag{3}$$

In addition, the sample variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})}$ in the feature space is defined as

$$\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})} = \frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)' \tag{4}$$

In order to perform the PCA in the feature space, the eigenvalue λ and the corresponding eigenvector $\mathbf{v}$ for $\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})}$ are computed by solving the following eigenvalue problem:

$$\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})}\mathbf{v} = \lambda\mathbf{v} \tag{5}$$

Since an eigenvector $\mathbf{v}$ lies in the span of $\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_n)$, the following holds.

$$\mathbf{v} = \sum_{l=1}^{n} \alpha_l \Phi(\mathbf{x}_l) \tag{6}$$

Multiplying both sides of Eq. (5) by $\Phi(\mathbf{x}_k)$ yields

$$\Phi(\mathbf{x}_k)'\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})}\mathbf{v} = \lambda\Phi(\mathbf{x}_k)'\mathbf{v}, k = 1, \ldots, n \tag{7}$$

Inserting Eqs. (4) and (6) into Eq. (7) yields

$$\mathbf{K}^2\boldsymbol{\alpha} = n\lambda\mathbf{K}\boldsymbol{\alpha} \tag{8}$$

where $\mathbf{K}$, which is called the gram matrix, is defined as

$$\mathbf{K} = \begin{bmatrix} \Phi(\mathbf{x}_1)'\Phi(\mathbf{x}_1) & \cdots & \Phi(\mathbf{x}_1)'\Phi(\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \Phi(\mathbf{x}_n)'\Phi(\mathbf{x}_1) & \cdots & \Phi(\mathbf{x}_n)'\Phi(\mathbf{x}_n) \end{bmatrix} \tag{9}$$

and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$. Solving a generalized eigenvalue problem in Eq. (8) is equivalent to solving the following eigenvalue problem:

$$\mathbf{K}\boldsymbol{\alpha} = n\lambda\boldsymbol{\alpha} \tag{10}$$

Since $\mathbf{K}$ is positive semidefinite, there are n nonnegative eigenvalues and the corresponding eigenvectors. Let $n\lambda_k$ be the kth largest eigenvalue and $\boldsymbol{\alpha}_k = (\alpha_{1k}, \ldots, \alpha_{nk})'$ be the corresponding eigenvector for $k = 1, \ldots, n$. Then, $\boldsymbol{\alpha}_k$ is normalized by normalizing the eigenvector $\mathbf{v}_k$ in the feature space as follows.

$$1 = \mathbf{v}'_k\mathbf{v}_k = \sum_{i,l=1}^{n} \alpha_{ik}\alpha_{lk}(\Phi(\mathbf{x}_i)'\Phi(\mathbf{x}_l)) = \boldsymbol{\alpha}'_k\mathbf{K}\boldsymbol{\alpha}_k = n\lambda_k(\boldsymbol{\alpha}'_k\boldsymbol{\alpha}_k)$$

In order to extract the principal component score t_{ik}, $\Phi(\mathbf{x}_i)$ is projected onto the eigenvector $\mathbf{v}_k$ as follows [see Eq. (6)].

$$t_{ik} = \mathbf{v}'_k\Phi(\mathbf{x}_i) = \sum_{l=1}^{n} \alpha_{lk}\Phi(\mathbf{x}_l)'\Phi(\mathbf{x}_i) = \sum_{l=1}^{n} \alpha_{lk}\mathbf{K}(\mathbf{x}_l, \mathbf{x}_i), k = 1, \ldots, n \tag{11}$$

where $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_i)$ is the (l, i)th element of the gram matrix, namely, $\Phi(\mathbf{x}_l)'\Phi(\mathbf{x}_i)$. As shown in Eq. (2), $\Phi(\mathbf{x}_l)'\Phi(\mathbf{x}_i)$ can be replaced by a kernel function $k(\mathbf{x}_l, \mathbf{x}_i)$ such as the one in Table 1, and therefore, the explicit form of $\Phi(\mathbf{x}_i)$ is not required.

In the PCA algorithm, it is assumed that the mapped data are centered as in Eq. (3). However, the explicit form of $\Phi(\cdot)$ is unknown in practice, and therefore, the mean of uncentered $\Phi(\mathbf{x}_i)$ cannot be calculated. This implies that the matrix $\mathbf{K}$ in Eq. (9), which is based on centered $\Phi(\mathbf{x}_i)$'s, cannot be constructed directly. Fortunately, however, $\mathbf{K}$ can be constructed as follows without actually centering $\Phi(\mathbf{x}_i)$'s (Schölkopf et al. 1998).

$$\mathbf{K} = \mathbf{K}_u - \mathbf{1}_n\mathbf{K}_u - \mathbf{K}_u\mathbf{1}_n + \mathbf{1}_n\mathbf{K}_u\mathbf{1}_n \tag{12}$$

where $\mathbf{K}_u$ is the gram matrix with uncentered $\Phi(\mathbf{x}_i)$'s and its (l, i)th element is $k(\mathbf{x}_l, \mathbf{x}_i)$, and $(\mathbf{1}_n)_{ij} = 1/n$.

The procedure of the kernel PCA to obtain the principal component scores requires to solve a similar eigenvalue problem as in Eq. (1) for the original PCA. The only difference is that the kernel PCA has to deal with a gram matrix $\mathbf{K}$ instead of a sample variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{x}}}$. The kernel PCA algorithm is applied to solve MCPD problems in Sect. 3.

3 Proposed method

A kernel PCA-based method is developed to deal with MCPD problems. The proposed method allows to capture nonlinear relationships among multiple performance characteristics in constructing a single aggregate performance measure. The proposed method proceeds according to the following steps.

(1) Calculate the SN ratio for each performance characteristic

Let y be a performance characteristic. Then, the Taguchi SN ratio is calculated for each y. Depending on the type of y, the expected loss and the corresponding SN ratio are defined in a different manner as in Table 2 (Yum et al. 2013).

For y of the NB type, it is assumed that the mean is adjusted to the target using an adjustment parameter, and therefore, the expected loss after adjustment is considered. Note from Table 2 that maximizing an SN ratio is equivalent to minimizing the corresponding expected loss (after adjustment when applicable).

(2) Standardize the SN ratio

For each performance characteristic, the estimated SN ratio is standardized as follows.

$$SN_{ij}^N = \frac{SN_{ij} - \overline{SN}_j}{SN_j^s} \tag{13}$$

where SN_{ij} is the SN ratio at the ith experimental run for the jth performance characteristic for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, and $\overline{SN}_j$ and SN_j^s are respectively the mean and standard deviation of SN_{ij}'s for the jth performance characteristic.

(3) Perform the kernel PCA on the standardized SN ratios

Let $\mathbf{x}_i = (SN_{i1}^N, SN_{i2}^N, \ldots, SN_{ip}^N)'$ for $i = 1, \ldots, n$. Then, the matrix $\mathbf{K}_u$ is constructed using a selected kernel function, and subsequently $\mathbf{K}$ is constructed using Eq. (12). Any commonly used kernel functions in Table 1 can be chosen. The eigenvector $\boldsymbol{\alpha}_k = (\alpha_{1k}, \ldots, \alpha_{nk})'$, $k = 1, \ldots, n$, for $\mathbf{K}$ is calculated by solving the eigenvalue problem in Eq. (10). Principal component scores t_{ik}, $k = 1, \ldots, n$, are then calculated using Eq. (11), and combined to have an aggregate performance measure (APM) as follows.

$$APM_i = \sum_{k=1}^{n} w_k t_{ik} \tag{14}$$

where APM_i is the value of the aggregate performance measure at the ith experimental run, w_k is the proportion of variance explained by the eigenvector corresponding to the kth largest eigenvalue for $\mathbf{K}$. That is,

$$w_k = \frac{n\lambda_k}{\sum_{k=1}^{n} n\lambda_k} = \frac{\lambda_k}{\sum_{k=1}^{n} \lambda_k}$$

where λ_k is the kth largest eigenvalue of $\hat{\boldsymbol{\Sigma}}_{\Phi(\mathbf{x})}$.

If some λ_k's are small (e.g., see the simulated example in Sect. 4), the contributions of the corresponding t_{ik}'s to the APM_i are negligible, and therefore, the corresponding t_{ik}'s can be ignored when calculating the APM_i for computational convenience. Also note that for y of the NB type, the adjustment is considered when the corresponding SN ratio is calculated before the kernel PCA is performed on the standardized SN ratios.

(4) Determine the optimal levels of design parameters

The APM_i can be statistically analyzed (e.g., analysis of variance) to identify those design parameters that have a significant effect on it. The optimal level of such a significant parameter is chosen as the one at which the APM_i is maximized. The level of a non-significant parameter may be selected based on such non-statistical factors as the ease of operation, cost for maintaining the level, etc. If such non-statistical information is not available or if the levels of a design parameter are indifferent with respect to those non-statistical factors, then the optimal level of a design parameter may be simply chosen as the one at which the APM is maximized (e.g., see the simulated example in Sect. 4).

Table 2 Expected loss and SN ratio for each type of performance characteristic

Type	Loss function	Expected loss	True SN ratio	Estimated SN ratio
Smaller-the-better (SB)	$L(y) = cy^2$	$L = E\left(cy^2\right)$ $= c\left(\mu^2 + \sigma^2\right)$	$-10\log\left[E\left(y^2\right)\right]$ $= -10\log\left(\mu^2 + \sigma^2\right)$	$-10\log\left(\frac{1}{q}\sum_{r=1}^{q} y_r^2\right)$
Larger-the-better (LB)	$L(y) = \frac{c}{y^2}$	$L = E\left(c/y^2\right)$ $\approx c\frac{1}{\mu^2}\left(1 + 3\frac{\sigma^2}{\mu^2}\right)$	$-10\log\left[E\left(\frac{1}{y^2}\right)\right]$ $= -10\log\left[\frac{1}{\mu^2}\left(1 + 3\frac{\sigma^2}{\mu^2}\right)\right]$	$-10\log\left(\frac{1}{q}\sum_{r=1}^{q}\frac{1}{y_r^2}\right)$
Nominal-the-best (NB)	$L(y) = c(y - \tau)^2$	$L =$ Expected loss after adjustment $= c\tau^2\frac{\sigma^2}{\mu^2}$	$-10\log\left(\frac{\sigma^2}{\mu^2}\right)$	$-10\log\left(\frac{s^2}{\bar{y}^2}\right)$

c: quality loss coefficient, q: number of noise conditions, μ: population mean of y, σ^2: population variance of y, s^2: sample variance of y, τ: target of y

4 Simulated example

The proposed method is applied to the simulated example for L_{18} orthogonal array experimental design. The homogeneous polynomial kernel is employed since it allows sufficient flexibility in capturing nonlinear structures of the data (Ben-Hur and Weston 2010). Also notice that the kernel PCA which employs the homogeneous polynomial kernel with the degree $d = 1$ reduces to the original PCA. Although the simulated examples in this section adopted the homogeneous polynomial kernel, it is also possible to employ other kernel functions.

Suppose that there are two performance characteristics y_j, $j = 1, 2$, of which y_1 is an NB type and y_2 is an SB type. Experimental data are generated using the L_{18} orthogonal array for six design parameters (A, B, C, D, E, and F) with three levels each. Interaction effects among design parameters are assumed to be negligible. Let $y_{1,ijklmn}$ be the observation of y_1 when the design parameters A, B, C, D, E, and F are at the levels i, j, k, l, m, and n (i.e., $A_iB_jC_kD_lE_mF_n$), respectively. Then, $y_{1,ijklmn}$ is generated from a normal distribution as follows.

$$y_{1,ijklmn} \sim N(\mu_{1,ijklmn}, \sigma^2_{1,ijklmn}) \tag{15}$$

where $\mu_{1,ijklmn}$ and $\sigma_{1,ijklmn}$ are the true mean and standard deviation of y_1, respectively, at $A_iB_jC_kD_lE_mF_n$. The Taguchi SN ratio for the NB type performance characteristic depends on σ/μ (see Table 2), and therefore, in generating simulated data for y_1, $(\sigma_1/\mu_1)_{ijklmn}$ and $\mu_{1,ijklmn}$ are specified, and $\sigma_{1,ijklmn}$ is determined using the following relationship.

$$\sigma_{1,ijklmn} = (\sigma_1/\mu_1)_{ijklmn} \times \mu_{1,ijklmn}$$

When the generated $y_{1,ijklmn}$ is negative, it is replaced with 0. Similarly, let $y_{2,ijklmn}$ be the observation of y_2 at $A_iB_jC_kD_lE_mF_n$. Then, $y_{2,ijklmn}$ is generated from a normal distribution as follows.

$$y_{2,ijklmn} \sim N(\mu_{2,ijklmn}, \sigma^2_{2,ijklmn}) \tag{16}$$

where $\mu_{2,ijklmn}$ and $\sigma_{2,ijklmn}$ are the true mean and standard deviation of $y_{2,ijklmn}$, respectively, at $A_iB_jC_kD_lE_mF_n$. When the generated $y_{2,ijklmn}$ is negative, it is replaced with 0.

The effects of design parameters on σ_1/μ_1, μ_1, σ_2 and μ_2 are assumed to be additive. For example, σ_1/μ_1 at $A_iB_jC_kD_lE_mF_n$ is assumed as

$$(\sigma_1/\mu_1)_{ijklmn} = m + a_i + b_j + c_k + d_l + e_m + f_n, \quad i, j, k, l, m, n = 1, 2, 3 \tag{17}$$

where m is the overall mean, and a_i, b_j, c_k, d_l, e_m, and f_n denote the effects of respective design parameters on $(\sigma_1/\mu_1)_{ijklmn}$ at $A_iB_jC_kD_lE_mF_n$, and it is assumed that $\sum_{i=1}^{3} a_i = \sum_{j=1}^{3} b_j = \sum_{k=1}^{3} c_k = \sum_{l=1}^{3} d_l = \sum_{m=1}^{3} e_m = \sum_{n=1}^{3} f_n = 0$. Similarly, μ_1, σ_2, and μ_2 at $A_iB_jC_kD_lE_mF_n$ are modeled in terms of the effects of design parameters as in Eq. (17). The effects of each design parameter on σ_1/μ_1 and μ_1 for y_1, and on σ_2 and μ_2 for y_2 are summarized in Tables 3a, b, respectively. '−' indicates that the corresponding design parameter has no effect. Then, the true means and standard deviations at each run are determined as shown in Table 4.

Using μ's and σ's in Table 4, the true SN ratios can be calculated based on the formulas in Table 2. Figure 2 shows a scatter plot of the true standardized SN ratios of the two performance characteristics. It is observed that the standardized SN ratios do not have an ellipsoidal structure, which means that the structure of the simulated data may not be explained well by the principal components of the original PCA.

Table 3 Effects of design parameters on (a) σ_1/μ_1 and μ_1 of y_1, (b) σ_2 and μ_2 of y_2

Effects of parameters	σ_1/μ_1			μ_1		
	1	2	3	1	2	3
(a)						
a	−0.1	0	0.1	5	0	−5
b	0	0.05	−0.05	0	−3	3
c	0.1	−0.1	0	–	–	–
d	0.05	0	−0.05	–	–	–
e	−0.1	0	0.1	−5	5	0
f	–	–	–	−2	0	2
m		0.6			60	

Effects of parameters	σ_2			μ_2		
	1	2	3	1	2	3
(b)						
a	–	–	–	2	−2	0
b	−0.05	0	0.05	–	–	–
c	−0.1	0.1	0	−2	0	2
d	0.2	0	−0.2	–	–	–
e	0	0.05	−0.05	2.5	0	−2.5
f	–	–	–	0	1.5	−1.5
m		0.6			10	

To evaluate the performance of the proposed method, a total of 10 simulated datasets are generated. One dataset consists of nine replicates of y_1 and of y_2 at each experimental run, which are generated respectively by Eqs. (15) and (16) using the parameter values in Table 4. This section illustrates the procedure of the proposed method with a dataset which is shown in Tables 5a, b for y_1 and y_2, respectively. First, the sample mean, sample standard deviation, SN ratio, and standardized SN ratio for each performance characteristic at each run are calculated as in Table 6, where the SN ratios are estimated based on formula in Table 2 and then standardized using Eq. (13).

Next, the kernel PCA is performed on the standardized SN ratios. The homogeneous polynomial kernel of degree d is chosen, and three cases where $d = 1, 2,$ and 3 are considered. When $d = 1$, the kernel PCA is reduced to the original PCA. When $d = 2$ and $d = 3$, the data structure can be explained by quadratic and cubic polynomial models, respectively. For each case, the matrix $\mathbf{K}_u$ is constructed using the homogeneous polynomial kernel in Table 1 and the matrix $\mathbf{K}$ is constructed using Eq. (12). Then, the eigenvalue problem for $\mathbf{K}$ is solved to find the eigenvalue λ_k and the corresponding eigenvector α_k for $k = 1, \ldots, n$. A total of eighteen eigenvalues and the corresponding eigenvectors are actually obtained for $\mathbf{K}$ since $\mathbf{K}$ is a 18×18 matrix. However, the eigenvalues which are close to zero (less than 10^{-10} in this example) and the corresponding eigenvectors are ignored for computational convenience since their contributions to the *APM* in Eq. (14) are negligible. Table 7 shows the computational results.

Now the standardized SN ratios are converted into the principal component scores using Eq. (11), which are then combined as the aggregate performance measure *APM* using Eq. (14).

| Table 4 True means and standard deviations of simulated data	Run	A	B	C	D	E	F	μ_1	μ_2	σ_1	σ_2
	1	1	1	1	1	1	1	58	12.5	31.90	0.65
	2	1	2	2	2	2	2	67	13.5	30.15	0.75
	3	1	3	3	3	3	3	70	10.0	35.00	0.40
	4	2	1	1	2	2	3	67	4.5	46.90	0.50
	5	2	2	2	3	3	1	55	5.5	33.00	0.45
	6	2	3	3	1	1	2	58	14.0	29.00	0.85
	7	3	1	2	1	3	3	57	6.0	42.75	0.80
	8	3	2	3	2	1	1	45	14.5	29.25	0.60
	9	3	3	1	3	2	2	63	9.5	44.10	0.40
	10	1	1	3	3	2	1	68	14.0	30.60	0.40
	11	1	2	1	1	3	2	62	9.0	49.60	0.65
	12	1	3	2	2	1	3	65	13.0	16.25	0.75
	13	2	1	2	3	1	2	55	12.0	19.25	0.45
	14	2	2	3	1	2	3	64	8.5	44.80	0.85
	15	2	3	1	2	3	1	61	3.5	45.75	0.50
	16	3	1	3	2	3	2	55	11.0	44.00	0.50
	17	3	2	1	3	1	3	49	9.0	34.30	0.30
	18	3	3	2	1	2	1	61	10.0	36.60	1.00
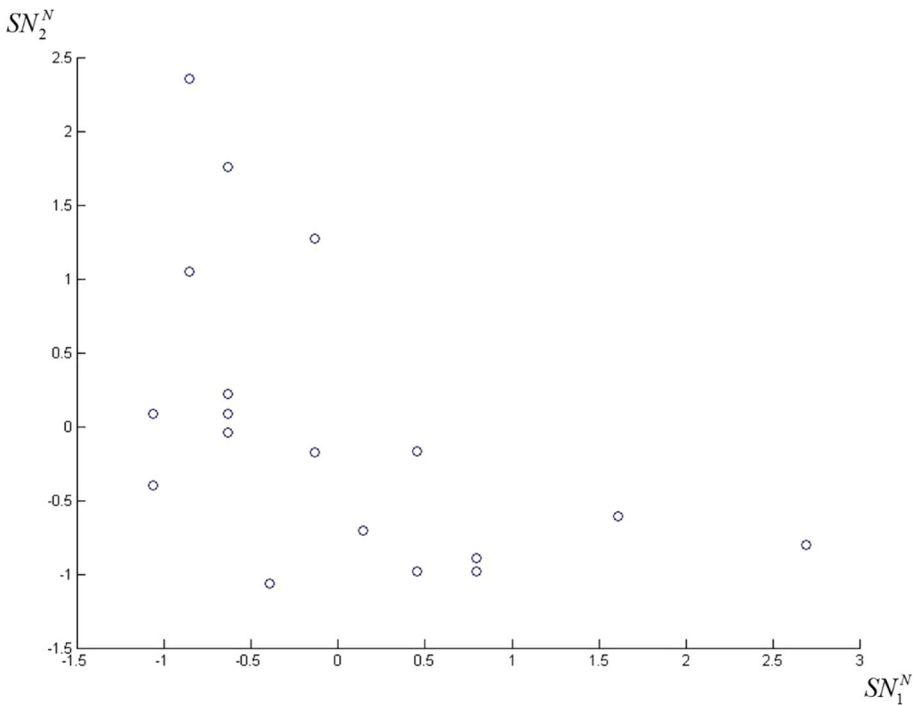
Fig. 2 Scatter plot of standardized SN ratios of simulated data

Table 5 Nine replicates of (a) y_1 and (b) y_2 in a dataset

Run	A	B	C	D	E	F	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9
(a)															
1	1	1	1	1	1	1	37.2	75.1	88.0	65.2	63.2	58.9	53.8	63.2	46.5
2	1	2	2	2	2	2	14.8	99.2	47.7	64.0	62.0	86.2	112.3	1.1	75.7
3	1	3	3	3	3	3	55.8	22.0	57.4	64.4	148.1	18.7	137.6	71.0	99.1
4	2	1	1	2	2	3	46.8	17.6	97.9	103.8	110.1	55.8	113.5	57.7	128.8
5	2	2	2	3	3	1	59.8	46.8	0.0	130.8	78.7	41.3	61.7	44.5	34.3
6	2	3	3	1	1	2	34.2	12.8	97.8	83.1	52.3	46.9	87.1	4.4	0.0
7	3	1	2	1	3	3	20.3	15.8	117.1	20.5	0.0	0.0	35.9	86.8	0.0
8	3	2	3	2	1	1	29.7	22.5	79.1	105.6	60.8	66.6	26.3	14.9	42.4
9	3	3	1	3	2	2	96.2	160.6	74.8	19.8	20.0	13.5	91.4	79.8	69.5
10	1	1	3	3	2	1	82.7	65.1	51.5	81.4	86.0	36.5	0.0	47.9	32.6
11	1	2	1	1	3	2	17.3	110.4	24.4	56.5	204.5	63.1	12.7	72.9	0.0
12	1	3	2	2	1	3	52.6	56.1	65.2	84.6	67.5	61.8	68.7	45.5	71.7
13	2	1	2	3	1	2	42.9	64.5	48.1	37.5	21.8	58.1	63.9	43.8	28.7
14	2	2	3	1	2	3	11.7	110.2	30.9	76.7	63.3	108.6	39.3	64.9	141.7
15	2	3	1	2	3	1	94.8	161.2	71.7	92.6	111.7	105.7	18.3	0.0	87.7
16	3	1	3	2	3	2	17.6	148.8	78.9	86.0	87.0	105.8	41.8	52.1	64.8
17	3	2	1	3	1	3	85.3	65.9	31.1	81.8	58.6	14.6	50.1	10.2	23.5
18	3	3	2	1	2	1	82.8	0.0	8.7	112.3	76.5	84.7	99.3	106.9	72.0
(b)															
1	1	1	1	1	1	1	13.1	12.0	12.3	12.3	12.1	12.2	12.8	12.3	12.6
2	1	2	2	2	2	2	13.2	13.5	14.2	12.2	14.5	13.9	14.6	14.3	12.6
3	1	3	3	3	3	3	9.9	9.8	9.4	10.6	9.9	9.9	10.4	9.9	9.9
4	2	1	1	2	2	3	4.7	4.8	4.9	4.8	4.9	4.3	3.6	5.5	4.2
5	2	2	2	3	3	1	6.2	6.2	5.9	5.8	5.6	5.4	5.3	5.2	5.2
6	2	3	3	1	1	2	15.4	12.3	13.8	13.7	14.8	14.2	13.2	15.1	15.3
7	3	1	2	1	3	3	6.2	6.1	6.3	5.6	7.1	4.5	7.2	5.4	5.3
8	3	2	3	2	1	1	14.6	14.7	14.8	14.9	13.7	14.9	14.9	14.9	14.6
9	3	3	1	3	2	2	9.6	9.2	9.5	9.8	8.9	9.3	9.0	10.0	9.7
10	1	1	3	3	2	1	14.5	13.9	14.2	14.2	14.4	14.1	14.2	13.3	14.5
11	1	2	1	1	3	2	8.7	8.6	7.9	9.1	9.1	9.2	9.4	8.8	9.7
12	1	3	2	2	1	3	12.7	12.1	12.7	12.5	13.1	13.3	12.2	13.2	12.6
13	2	1	2	3	1	2	12.4	11.7	12.0	11.8	11.9	12.2	11.7	11.4	11.6
14	2	2	3	1	2	3	7.4	8.9	8.7	9.1	7.8	8.8	8.0	9.0	9.0
15	2	3	1	2	3	1	3.6	4.5	3.0	3.4	3.5	3.6	3.9	4.2	3.2
16	3	1	3	2	3	2	11.1	10.0	10.2	11.1	11.0	10.7	10.2	10.8	11.5
17	3	2	1	3	1	3	9.1	8.7	9.0	9.3	9.2	8.3	9.1	9.2	9.0
18	3	3	2	1	2	1	8.9	10.2	10.4	9.5	9.7	10.0	10.4	10.4	10.6

Table 6 Sample means, sample standard deviations, estimated SN ratios, and standardized SN ratios of simulated data

Run	A	B	C	D	E	F	$\bar{y}_1$	s_1	SN_1	SN_1^N	$\bar{y}_2$	s_2	SN_2	SN_2^N
1	1	1	1	1	1	1	61.24	14.94	12.25	1.83	12.40	0.35	−21.87	−0.72
2	1	2	2	2	2	2	62.55	36.79	4.61	−0.16	13.67	0.86	−22.73	−0.97
3	1	3	3	3	3	3	74.90	45.59	4.31	−0.23	9.98	0.34	−19.99	−0.16
4	2	1	1	2	2	3	81.34	37.68	6.68	0.38	4.64	0.53	−13.39	1.78
5	2	2	2	3	3	1	55.33	35.68	3.81	−0.36	5.65	0.39	−15.07	1.28
6	2	3	3	1	1	2	46.53	36.88	2.02	−0.83	14.20	1.08	−23.07	−1.07
7	3	1	2	1	3	3	32.94	41.61	−2.03	−1.88	5.98	0.86	−15.62	1.12
8	3	2	3	2	1	1	49.78	30.29	4.31	−0.23	14.67	0.39	−23.33	−1.14
9	3	3	1	3	2	2	69.53	47.02	3.40	−0.47	9.45	0.39	−19.51	−0.02
10	1	1	3	3	2	1	53.76	28.38	5.55	0.09	14.15	0.37	−23.02	−1.05
11	1	2	1	1	3	2	62.42	63.72	−0.18	−1.40	8.93	0.52	−19.04	0.12
12	1	3	2	2	1	3	63.76	11.50	14.87	2.51	12.71	0.44	−22.09	−0.78
13	2	1	2	3	1	2	45.49	14.91	9.69	1.16	11.84	0.33	−21.47	−0.60
14	2	2	3	1	2	3	71.91	42.13	4.64	−0.15	8.52	0.63	−18.63	0.24
15	2	3	1	2	3	1	82.64	48.63	4.61	−0.16	3.64	0.48	−11.30	2.39
16	3	1	3	2	3	2	75.87	38.24	5.95	0.19	10.73	0.51	−20.62	−0.35
17	3	2	1	3	1	3	46.78	28.26	4.38	−0.22	8.99	0.31	−19.08	0.10
18	3	3	2	1	2	1	71.47	40.44	4.95	−0.07	10.01	0.57	−20.02	−0.17

Table 7 Gram matrix **K** and the solution of eigenvalue problem with respect to the degree of the homogeneous polynomial kernel

d	**K**	λ	α
1	$\begin{bmatrix} 3.8642 & \cdots & -0.0033 \\ \vdots & \ddots & \vdots \\ -0.0033 & \cdots & 0.0343 \end{bmatrix}_{18\times18}$	1.2918 0.7082	$(0.0774, \ldots, 0.0031)'_{1\times18}$ $(0.0619, \ldots, 0.0134)'_{1\times18}$
2	$\begin{bmatrix} 7.8450 & \cdots & -2.4857 \\ \vdots & \ddots & \vdots \\ -2.4857 & \cdots & 2.1167 \end{bmatrix}_{18\times18}$	3.8473 1.8845 0.2754	$(0.0404, \ldots, 0.0130)'_{1\times18}$ $(0.0007, \ldots, 0.0332)'_{1\times18}$ $(0.0115, \ldots, 0.0392)'_{1\times18}$
3	$\begin{bmatrix} 47.7877 & \cdots & -4.1428 \\ \vdots & \ddots & \vdots \\ -4.1428 & \cdots & 1.6270 \end{bmatrix}_{18\times18}$	27.3130 11.5765 0.9328 0.4548	$(-0.0140, \ldots, 0.0013)'_{1\times18}$ $(0.0012, \ldots, 0.0051)'_{1\times18}$ $(0.0195, \ldots, 0.0140)'_{1\times18}$ $(-0.0156, \ldots, 0.0131)'_{1\times18}$

Table 8 shows the resulting principal component scores and *APM*. The ANOVA is performed on *APM* to identify statistically significant parameters. A design parameter whose F ratio is larger than two is considered to be statistically significant (Phadke 1989). When $d = 1$ or 3, the design parameter E is identified as significant, while the design parameters A, C, and E are identified as significant when $d = 2$. To determine the optimal levels of design parameters, the main-effect plots are constructed. Figure 3 shows the main-effect plots for *APM* with respect to the degree d of the homogeneous polynomial kernel. It is observed that the optimal levels of design parameters for $d = 1, 2$, and 3 are $A_1B_1C_2D_2E_1F_1$, $A_1B_1C_2D_1E_1F_3$, and $A_1B_3C_2D_2E_1F_3$, respectively.

The SN ratio for each performance characteristic at the optimal levels is calculated based on the formula in Table 2 to see whether the proposed method performs effectively for each performance characteristic. For example, for $d = 1$, the SN ratio for each performance characteristic is calculated as follows.

$$\frac{\sigma_1}{\mu_1} = 0.6 - 0.1 - 0.1 - 0.1 = 0.3 \text{ from Table 3(a),}$$

$$\mu_2 = 10 + 2 + 2.5 = 14.5 \text{ and } \sigma_2 = 0.6 - 0.05 + 0.1 = 0.65 \text{ from Table 3(b),}$$

$$SN_1 = -10 \log \left(\frac{\sigma_1^2}{\mu_1^2} \right) = -10 \log \left(0.3^2 \right) \approx 10.4576,$$

$$SN_2 = -10 \log \left[E \left(y^2 \right) \right] = -10 \log \left(\mu_2^2 + \sigma_2^2 \right) = -10 \log \left(14.5^2 + 0.65^2 \right) \approx -23.2361.$$

Similarly, SN_1 and SN_2 at the optimal levels for $d = 2$ are calculated as 9.1186 and -22.2974, respectively, and for $d = 3$ as 12.0412 and -22.2933, respectively.

The proposed method is then applied to the remaining nine datasets using the same procedures as above. Table 9 shows separate SN ratios for each performance characteristic at the optimal conditions. The largest SN ratio is italicized in each trial, and the kernel PCA produces largest SN ratios for both performance characteristics at trials 1, 2, 4, 6, 7, 9, and 10, while the original PCA never produces largest SN ratios simultaneously for both characteristics.

The proposed method is also applied to the case of L_9 orthogonal array simulated experimental design under the same assumptions on the performance characteristics as for the L_{18} case. Table 10 shows the results for the L_9 case. The largest SN ratio is italicized in each trial. It is observed that the kernel PCA-based method usually leads to larger SN ratio for each characteristic than those from the original PCA-based method.

5 Low-pressure cold spray process example

The proposed method is also applied to the analysis of the low-pressure cold spray (LPCS) process data (Goyal et al. 2013). The experiments were performed to determine the optimal levels of LPCS parameters with respect to three performance characteristics, namely, coating thickness (y_1) which is the LB type, coating density (y_2) which is the LB type, and surface roughness (y_3) which is the SB type performance characteristic. Five LPCS parameters were considered using the L_{18} orthogonal array.

Figure 4 shows a scatter plot of standardized SN ratios of three performance characteristics. The estimated SN ratios for each type are standardized using Eq. (13). It is observed that the standardized SN ratios do not have an ellipsoidal structure, which means that the structure of the LPCS data may not be explained well enough by the principal components of the original PCA.

Table 8 Principal component scores t and APM of the simulated dataset with respect to the degree of the homogeneous polynomial kernel

| d	1			2				3				
λ	1.292	0.708		3.847	1.885	0.275		27.31	11.58	0.933	0.455	
w	0.646	0.354		0.640	0.314	0.046		0.678	0.287	0.023	0.011	
#	t_1	t_2	APM	t_1	t_2	t_3	APM	t_1	t_2	t_3	t_4	APM
1	1.801	0.789	1.442	2.800	0.023	−0.057	1.798	6.900	0.241	−0.328	−0.128	4.739
2	0.574	−0.796	0.089	−1.132	0.255	0.035	−0.644	−0.542	1.960	0.009	−0.019	0.196
3	−0.050	−0.280	−0.132	−0.876	1.128	−0.137	−0.214	−0.651	1.069	−0.237	0.088	−0.139
4	−0.985	1.526	−0.096	−1.765	−1.842	0.857	−1.669	−1.259	−4.454	−2.138	1.309	−2.168
5	−1.166	0.650	−0.523	−0.721	−0.571	−0.623	−0.669	−1.399	−0.993	0.639	−0.182	−1.222
6	0.167	−1.342	−0.367	−1.117	0.116	1.278	−0.620	−0.622	2.297	0.640	−1.849	0.233
7	−2.125	−0.539	−1.563	3.408	−0.882	−0.889	1.865	−10.75	1.100	2.521	1.210	−6.904
8	0.644	−0.974	0.071	−1.250	−0.093	0.211	−0.820	−0.482	2.542	0.251	−0.187	0.408
9	−0.318	−0.350	−0.329	−0.710	1.129	−0.086	−0.104	−0.729	1.080	−0.282	0.062	−0.190
10	0.806	−0.683	0.279	−1.004	0.044	−0.259	−0.641	−0.420	2.213	−0.335	0.199	0.346
11	−1.075	−0.908	−1.016	0.896	0.884	0.594	0.878	−3.103	1.513	−1.211	−0.648	−1.705
12	2.327	1.227	1.937	5.759	−0.527	0.682	3.554	17.29	−1.406	1.499	0.582	11.362
13	1.244	0.401	0.946	0.690	0.519	−0.342	0.589	1.545	0.950	−0.612	−0.082	1.305
14	−0.271	0.062	−0.153	−0.858	1.089	−0.241	−0.219	−0.660	1.052	−0.224	0.110	−0.149
15	−1.802	1.579	−0.605	−1.549	−4.548	−0.346	−2.435	−3.193	−12.40	0.552	−0.799	−5.725
16	0.382	−0.110	0.208	−0.833	1.017	−0.268	−0.227	−0.615	1.103	−0.274	0.114	−0.105
17	−0.226	−0.080	−0.175	−0.838	1.133	−0.215	−0.191	−0.661	1.065	−0.236	0.111	−0.146
18	0.073	−0.170	−0.013	−0.901	1.126	−0.194	−0.233	−0.647	1.069	−0.235	0.108	−0.136
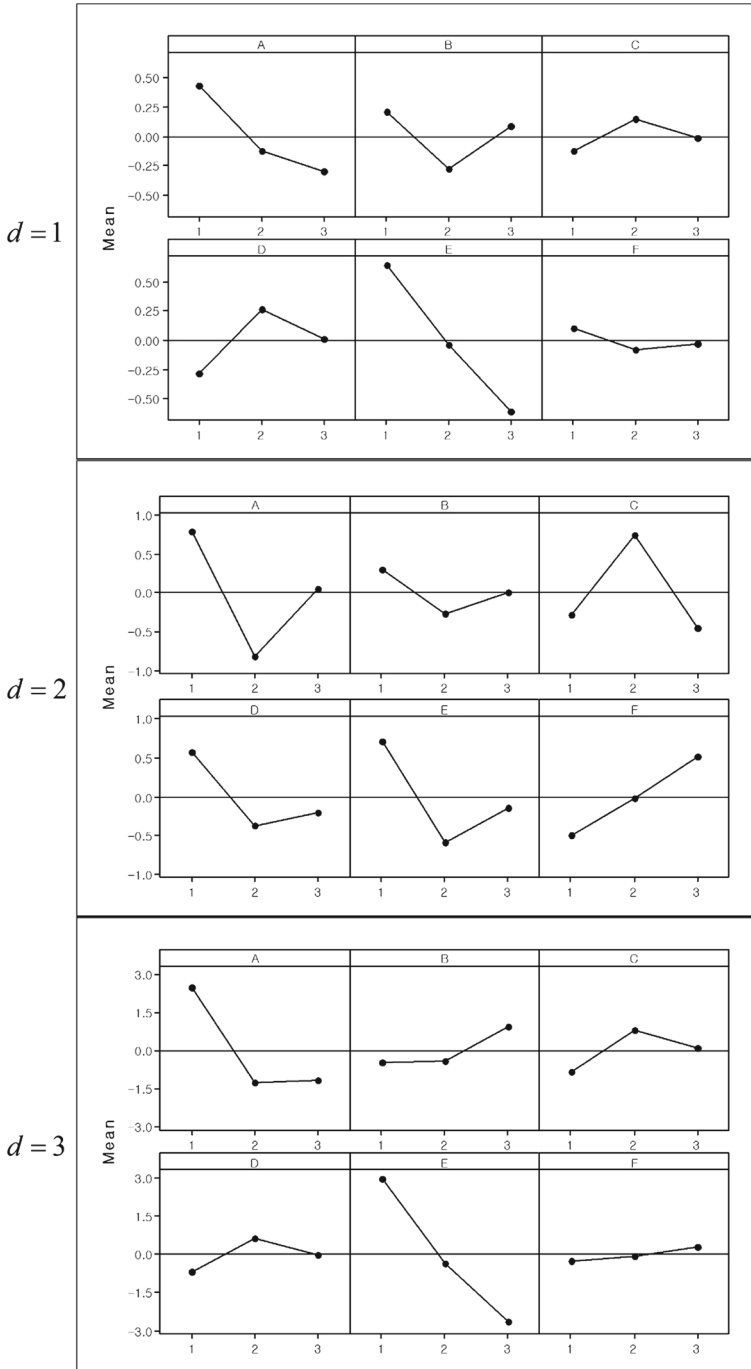
Fig. 3 The means of *APM* at each level of design parameters with respect to the degree of the homogeneous polynomial kernel

Table 9 Separate SN ratios for each performance characteristic at the optimal conditions with respect to the degree of the homogeneous polynomial kernel for L_{18} case

Trials	$d = 1$		
	Optimal levels	SN_1	SN_2
1	$A_1B_1C_2D_2E_1F_1$	10.4576	−23.2361
2	$A_1B_3C_2D_2E_1F_2$	*12.0412*	−24.0619
3	$A_1B_3C_2D_3E_1F_2$	*13.9794*	−24.0875
4	$A_1B_1C_2D_2E_1F_2$	10.4576	−24.0896
5	$A_1B_3C_3D_3E_1F_2$	*10.4576*	−25.1082
6	$A_1B_3C_2D_2E_1F_2$	*12.0412*	−24.0919
7	$A_1B_3C_2D_2E_1F_2$	12.0412	−24.0919
8	$A_1B_1C_2D_2E_1F_3$	10.4576	−22.2897
9	$A_1B_3C_2D_2E_1F_3$	12.0412	−22.2933
10	$A_1B_3C_3D_3E_1F_2$	10.4576	−25.1082

Trials	$d = 2$		
	Optimal levels	SN_1	SN_2
1	$A_1B_1C_2D_1E_1F_3$	9.1186	−22.2974
2	$A_1B_3C_2D_2E_1F_3$	*12.0412*	−22.2933
3	$A_1B_3C_2D_2E_1F_3$	12.0412	−22.2933
4	$A_1B_3C_2D_2E_1F_3$	*12.0412*	−22.2933
5	$A_2B_3C_2D_2E_1F_3$	9.1186	*−19.1149*
6	$A_1B_3C_2D_2E_1F_3$	*12.0412*	−22.2933
7	$A_1B_3C_2D_3E_1F_3$	13.9794	−22.2866
8	$A_1B_3C_2D_2E_1F_3$	*12.0412*	−22.2933
9	$A_1B_3C_2D_3E_1F_3$	*13.9794*	−22.2866
10	$A_1B_3C_2D_2E_1F_3$	*12.0412*	−22.2933

Trials	$d = 3$		
	Optimal levels	SN_1	SN_2
1	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*
2	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*
3	$A_1B_3C_2D_2E_1F_3$	12.0412	*−22.2933*
4	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*
5	$A_1B_2C_2D_3E_1F_2$	*10.4576*	−24.0866
6	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*
7	$A_1B_3C_2D_2E_1F_3$	12.0412	−22.2933
8	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*
9	$A_1B_3C_2D_2E_1F_3$	12.0412	−22.2933
10	$A_1B_3C_2D_2E_1F_3$	*12.0412*	*−22.2933*

Table 10 Separate SN ratios for each performance characteristic at the optimal conditions with respect to the degree of the homogeneous polynomial kernel for L_9 case

| Trials	$d = 1$		
	Optimal levels	SN_1	SN_2
1	$A_2B_2C_2D_1$	12.0412	−12.1722
2	$A_3B_2C_3D_2$	12.0412	−15.5823
3	$A_2B_2C_3D_1$	12.0412	−12.1378
4	$A_2B_2C_2D_1$	12.0412	−12.1722
5	$A_3B_2C_3D_1$	16.4782	−16.9338
6	$A_2B_2C_2D_1$	12.0412	−12.1722
7	$A_2B_2C_3D_1$	12.0412	−12.1378
8	$A_2B_2C_3D_1$	12.0412	−12.1378
9	$A_1B_1C_1D_1$	7.9588	−20.0088
10	$A_2B_2C_2D_1$	12.0412	−12.1722
Trials	$d = 2$		
	Optimal levels	SN_1	SN_2
1	$A_2B_2C_3D_1$	12.0412	−12.1378
2	$A_2B_2C_3D_1$	12.0412	−12.1378
3	$A_2B_2C_1D_1$	12.0412	−12.1085
4	$A_2B_2C_3D_1$	12.0412	−12.1378
5	$A_2B_2C_3D_1$	12.0412	−12.1378
6	$A_2B_2C_3D_1$	12.0412	−12.1378
7	$A_2B_2C_3D_2$	9.1186	−9.6190
8	$A_2B_2C_3D_1$	12.0412	−12.1378
9	$A_2B_2C_3D_2$	9.1186	−9.6190
10	$A_2B_2C_3D_1$	12.0412	−12.1378
Trials	$d = 3$		
	Optimal levels	SN_1	SN_2
1	$A_2B_2C_3D_1$	12.0412	−12.1378
2	$A_3B_2C_3D_1$	16.4782	−16.9338
3	$A_2B_2C_3D_1$	12.0412	−12.1378
4	$A_2B_2C_3D_1$	12.0412	−12.1378
5	$A_2B_2C_3D_1$	12.0412	−12.1378
6	$A_2B_2C_3D_1$	12.0412	−12.1378
7	$A_2B_2C_3D_1$	12.0412	−12.1378
8	$A_2B_2C_3D_1$	12.0412	−12.1378
9	$A_3B_2C_2D_3$	9.1186	−13.9950
10	$A_2B_2C_3D_1$	12.0412	−12.1378
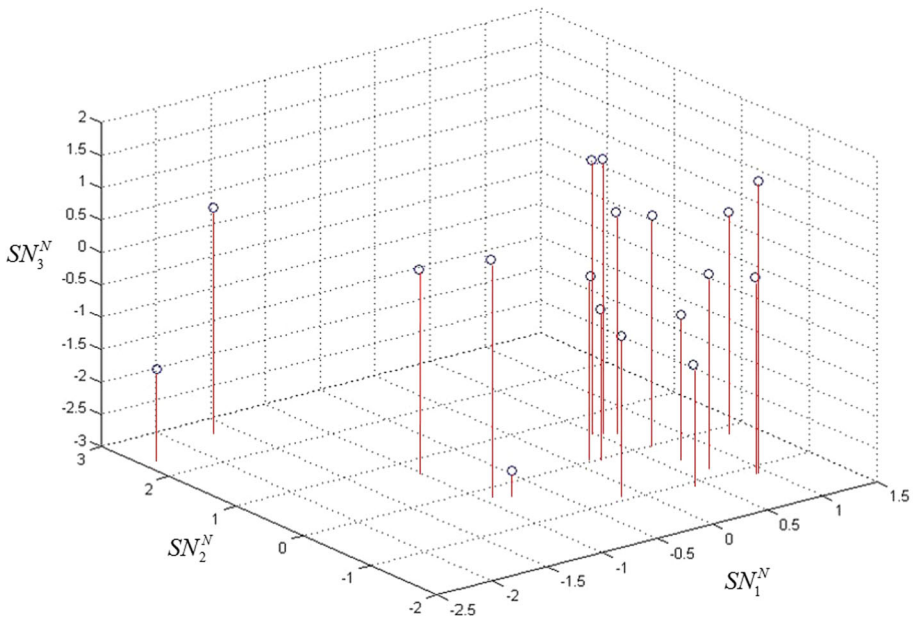
Fig. 4 Scatter plot of standardized SN ratios of LPCS data

The homogeneous polynomial kernel is chosen, and the optimal levels of the design parameters are determined using the proposed method as illustrated in Sect. 4. Table 11 shows the eigenvalue λ_k and the corresponding eigenvector α_k for the solution of the gram matrix **K** in the LPCS process example. The eigenvalues which are close to zero (less than 10^{-10} in this example) and the corresponding eigenvectors are ignored for computational convenience.

Then, the principal component scores are calculated using Eq. (11), and they are combined as the aggregate performance measure *APM* using Eq. (14). Table 12 shows the *APM* values for each degree d of the homogeneous polynomial kernel.

The optimal levels of design parameters are determined based on the main-effect plots of the *APM* as in Fig. 3. Since it is in general impossible for a reported case to evaluate the performance of the method in terms of the true parameter values, the SN ratio for each performance characteristic at the optimal condition is predicted instead. To predict the SN ratio for each performance characteristic, an additive model is used. For example, suppose that the optimal levels of the design parameters A $\sim$ E are identified as $A_i B_j C_k D_l E_m$, and the design parameters A and B have a significant effect on the SN ratio of a performance characteristic. Then, the SN ratio of this performance characteristic at $A_i B_j C_k D_l E_m$ is predicted as follows.

$$\widehat{SN}_{opt} = \overline{SN} + \left(\overline{SN}_{A_i} - \overline{SN}\right) + \left(\overline{SN}_{B_j} - \overline{SN}\right)$$

where $\overline{SN}$ is an overall mean of the SN ratios, $\overline{SN}_{A_i}$ and $\overline{SN}_{B_j}$ are the means of the SN ratios when the design parameters A and B are at levels i and j, respectively.

The significance of the design parameters is identified by checking the F ratio from the results of the analysis of variance (ANOVA). In general, if the F ratio is less than two, the corresponding effect is considered insignificant. The F ratio larger than two indicates that

Table 11 Gram matrix **K** and the solution of eigenvalue problem with respect to the degree of the homogeneous polynomial kernel for LPCS process example

d	**K**	λ	α
1	$\begin{bmatrix} 7.2301 & \cdots & -1.1499 \\ \vdots & \ddots & \vdots \\ -1.1499 & \cdots & 9.5927 \end{bmatrix}_{18 \times 18}$	1.7669	$(0.0337, \ldots, 0.0831)'_{1 \times 18}$
		1.0478	$(0.1307, \ldots, -0.0857)'_{1 \times 18}$
		0.1852	$(0.0051, \ldots, -0.0074)'_{1 \times 18}$
2	$\begin{bmatrix} 39.7223 & \cdots & -17.8980 \\ \vdots & \ddots & \vdots \\ -17.8980 & \cdots & 66.1312 \end{bmatrix}_{18 \times 18}$	10.2919	$(0.0046, \ldots, -0.0296)'_{1 \times 18}$
		3.7447	$(-0.0716, \ldots, 0.0772)'_{1 \times 18}$
		2.0132	$(0.1091, \ldots, 0.0829)'_{1 \times 18}$
		0.2258	$(0.0150, \ldots, -0.0377)'_{1 \times 18}$
		0.1610	$(0.1072, \ldots, -0.0235)'_{1 \times 18}$
		0.0500	$(-0.0798, \ldots, 0.0926)'_{1 \times 18}$
3	$\begin{bmatrix} 352.1451 & \cdots & -84.3581 \\ \vdots & \ddots & \vdots \\ -84.3581 & \cdots & 742.8457 \end{bmatrix}_{18 \times 18}$	137.6336	$(0.7938, \ldots, -15.5719)'_{1 \times 18}$
		34.8204	$(-6.7843, \ldots, 21.9270)'_{1 \times 18}$
		20.6261	$(17.4727, \ldots, 4.4008)'_{1 \times 18}$
		1.7020	$(-0.1060, \ldots, 0.3540)'_{1 \times 18}$
		0.4287	$(-0.3356, \ldots, 0.2694)'_{1 \times 18}$
		0.2367	$(0.2369, \ldots, 0.0009)'_{1 \times 18}$
		0.1282	$(0.0854, \ldots, 0.0055)'_{1 \times 18}$
		0.0816	$(-0.0288, \ldots, 0.0142)'_{1 \times 18}$
		0.0346	$(0.0530, \ldots, -0.0537)'_{1 \times 18}$
		0.0120	$(0.0453, \ldots, -0.0090)'_{1 \times 18}$

the corresponding effect is not quite small, whereas the F ratio larger than four means that the corresponding effect is statistically significant (Phadke 1989). In this paper, a design parameter whose F ratio is larger than two is considered as statistically significant.

Table 13 shows the predicted SN ratio for each performance characteristic at the optimal condition. The ANOVA results indicate that all design parameters A $\sim$ E have a significant effect on each SN ratio, and therefore, all of their effects are reflected to predict the SN ratios. The largest predicted SN ratio is italicized for each degree d of the homogeneous polynomial kernel.

In Table 13, the original PCA ($d = 1$) produces a larger predicted SN ratio for the first performance characteristic, while the kernel PCA ($d = 2, 3$) yields larger predicted SN ratios for the second and third performance characteristics. Although neither the original PCA nor the kernel PCA produces consistently larger SN ratios for all performance characteristics, the kernel PCA-based method would be preferred unless the importance of the first performance characteristic is substantially higher than the others.

Table 12 *APM* of the LPCS process data with respect to the degree of the homogeneous polynomial kernel

Run	A	B	C	D	E	$d = 1$	$d = 2$	$d = 3$
1	1	1	1	1	1	0.2304	−0.0827	1.1914
2	1	1	2	2	2	0.8875	1.2278	1.9143
3	1	1	3	3	3	0.6076	0.0956	3.2506
4	1	2	1	1	2	0.6469	1.0765	1.9611
5	1	2	2	2	3	0.0756	1.0402	1.9668
6	1	2	3	3	1	−0.8102	0.5379	1.3514
7	1	3	1	2	1	−1.9746	−7.6404	−32.7749
8	1	3	2	3	2	−0.2133	0.9190	1.9352
9	1	3	3	1	3	−0.2714	1.0050	1.8931
10	2	1	1	3	3	0.9675	0.1226	3.3394
11	2	1	2	1	1	0.4700	1.1000	1.9612
12	2	1	3	2	2	0.7226	0.4345	2.6689
13	2	2	1	2	3	0.2265	1.0650	1.9706
14	2	2	2	3	1	−0.4170	1.0753	1.9715
15	2	2	3	1	2	0.7606	−1.7962	7.7764
16	2	3	1	3	2	−0.0136	1.0137	1.9709
17	2	3	2	1	3	0.2271	0.6794	2.2346
18	2	3	3	2	1	−2.1221	−1.8732	−6.5825

Table 13 Predicted SN ratios for each performance characteristic at the optimal levels of LPCS parameters with respect to the degree of the homogeneous polynomial kernel

Parameter	Optimal levels	$\widehat{SN}_1$	$\widehat{SN}_2$	$\widehat{SN}_3$
$d = 1$	$A_2B_1C_2D_1E_2$	*38.6878*	63.6506	−18.8717
$d = 2$	$A_2B_2C_2D_3E_3$	36.4828	*70.4489*	−16.5267
$d = 3$	$A_2B_2C_2D_1E_2$	36.9778	66.6272	*−17.5367*

6 Conclusion

A kernel PCA-based method is developed to deal with the MCPD problems. The proposed method allows to capture possible nonlinear relationships among multiple performance characteristics in constructing a single aggregate performance measure, and therefore, is more flexible than the existing original PCA-based methods that only allow linear feature extraction.

Computational results for the problems with simulated data indicate that the kernel PCA-based method generally performs better than the original PCA-based method. Application of the proposed method to a real dataset also shows its potential for better performance than the original PCA.

The performance of the proposed method depends on the choice of a kernel function and its parameters. However, there are no universally accepted guideline on how to choose the proper kernel function (Lampert 2009). Instead, there are some heuristic approaches (e.g., see Hsu et al. 2003) for determining the kernel parameters using the so called cross validation. However, those approaches are usually applicable for large datasets, and may

not be adequate for the experimental data in MCPD problems. Further research needs to be conducted to provide guidelines for selecting kernel parameters as well as kernel functions in solving various MCPD problems.

The proposed method was applied to the simulated and real datasets. For a more thorough evaluation, it is desired that the proposed method be tested and compared with other existing methods for the MCPD problems with various experimental designs and a diverse type and number of performance characteristics.

References

Antony, J., Anand, R. B., Kumar, M., & Tiwari, M. K. (2006). Multiple response optimization using Taguchi methodology and neuro-fuzzy based model. *Journal of Manufacturing Technology Management*, *17*(7), 908–925.

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. In O. Carugo & F. Eisenhaber (Eds.), *Data mining techniques for the life sciences* (pp. 223–239). New York: Human Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Datta, S., Nandi, G., Bandyopadhyay, A., & Pal, P. K. (2009). Application of PCA-based hybrid Taguchi method for correlated multicriteria optimization of submerged arc weld: A case study. *International Journal of Advanced Manufacturing Technology*, *45*(3–4), 276–286.

Forouraghi, B. (2000). A genetic algorithm for multiobjective robust design. *Applied Intelligence*, *12*(3), 151–161.

Goyal, T., Walia, R. S., & Sidhu, T. S. (2013). Multi-response optimization of low-pressure cold-sprayed coatings through Taguchi method and utility concept. *The International Journal of Advanced Manufacturing Technology*, *64*(5–8), 903–914.

Hsu, C. H., Chang, C. C., & Lin, C. J. (2003). *A practical guide to support vector classification* (pp. 1–16).

Hsu, C. M. (2001). Solving multi-response problems through neural networks and principal component analysis. *Journal of the Chinese Institute of Industrial Engineers*, *18*(5), 47–54.

Huang, C. C., & Tang, T. T. (2006). Parameter optimization in melt spinning by neural networks and genetic algorithms. *International Journal of Advanced Manufacturing Technology*, *27*(11–12), 1113–1118.

Jean, M. D., & Wang, J. T. (2006). Using a principal components analysis for developing a robust design of electron beam welding. *International Journal of Advanced Manufacturing Technology*, *28*(9–10), 882–889.

Jeyapaul, R., Shahabudeen, P., & Krishnaiah, K. (2006). Simultaneous optimization of multi-response problems in the Taguchi method using genetic algorithm. *International Journal of Advanced Manufacturing Technology*, *30*(9–10), 870–878.

Jolliffe, I. T. (2002). *Principal component analysis*. New York: Springer.

Jung, J. H., & Kwon, W. T. (2010). Optimization of EDM process for multiple performance characteristics using Taguchi method and grey relational analysis. *Journal of Mechanical Science and Technology*, *24*(5), 1083–1090.

Lampert, C. H. (2009). *Kernel methods in computer vision*. Hanover, MA: Now Publishers Inc.

Liao, H. C. (2006). Multi-response optimization using weighted principal component. *International Journal of Advanced Manufacturing Technology*, *27*(7–8), 720–725.

Lin, C. L. (2004). Use of the Taguchi method and grey relational analysis to optimize turning operations with multiple performance characteristics. *Materials and Manufacturing Processes*, *19*(2), 209–220.

Lin, J. L., & Lin, C. L. (2002). The use of the orthogonal array with grey relational analysis to optimize the electrical discharge machining process with multiple performance characteristics. *International Journal of Machine Tools & Manufacture*, *42*(2), 237–244.

Lin, J. L., & Lin, C. L. (2005). The use of grey-fuzzy logic for the optimization of the manufacturing process. *Journal of Materials Processing Technology*, *160*(1), 9–14.

Lin, J. L., & Tarng, Y. S. (1998). Optimization of the multi-response process by the Taguchi method with grey relational analysis. *Journal of Grey System*, *4*(4), 355–370.

Lin, J. L., Wang, K. S., Yan, B. H., & Tarng, Y. S. (2000). Optimization of the electrical discharge machining process based on the Taguchi method with fuzzy logics. *Journal of Materials Processing Technology*, *102*(1), 48–55.

Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, *12*(2), 181–202.

Phadke, M. S. (1989). *Quality engineering using robust design*. New Jersey: Prentice Hall.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels*. Cambridge, MA: MIT Press.

Sharma, V., Chattopadhyaya, S., & Hloch, S. (2011). Multi response optimization of process parameters based on Taguchi-Fuzzy model for coal cutting by water jet technology. *International Journal of Advanced Manufacturing Technology*, *56*(9–12), 1019–1025.

Sibalija, T. V., & Majstorovic, V. D. (2009). Multi-response optimisation of thermosonic copper wire-bonding process with correlated responses. *International Journal of Advanced Manufacturing Technology*, *42*(3–4), 363–371.

Sibalija, T. V., & Majstorovic, V. D. (2012). An integrated approach to optimise parameter design of multi-response processes based on Taguchi method and artificial intelligence. *Journal of Intelligent Manufacturing*, *23*(5), 1511–1528.

Su, C. T., & Tong, L. I. (1997). Multi-response robust design by principal component analysis. *Total Quality Management*, *8*(6), 409–416.

Sukthomya, W., & Tannock, J. D. T. (2005). Taguchi experiemental design for manufacturing process optimisation using historical data and a neural network process model. *International Journal of Quality & Reliability Management*, *22*(5), 485–502.

Taguchi, G., & Yokoyama, Y. (1993). *Taguchi methods: Design of experiments, quality engineering series vol 4*. Michigan: ASI Press.

Tarng, Y. S., Juang, S. C., & Chang, C. H. (2002). The use of grey-based Taguchi methods to determine submerged arc welding process parameters in hardfacing. *Journal of Materials Processing Technology*, *128*(1), 1–6.

Tong, L. I., & Su, C. T. (1997). Optimizing multi-response problems in the Taguchi method by fuzzy multiple attribute decision making. *Quality and Reliability Engineering International*, *13*(1), 25–34.

Tsao, C. C., & Hocheng, H. (2008). Evaluation of thrust force and surface roughness in drilling composite material using Taguchi analysis and neural network. *Journal of Materials Processing Technology*, *203*(1), 342–348.

Tzeng, C. J., Lin, Y. H., Yang, Y. K., & Jeng, M. C. (2009). Optimization of turning operations with multiple performance characteristics using the Taguchi method and grey relational analysis. *Journal of Materials Processing Technology*, *209*(6), 2753–2759.

Wang, Z., Zhu, L. I., & Wu, J. H. (1996). Grey relational analysis of correlation of errors in measurement. *Journal of Grey System*, *8*(1), 73–78.

Yang, Y. S., Shih, C. Y., & Fung, R. F. (2014). Multi-objective optimization of the light guide rod by using the combined Taguchi method and grey relational approach. *Journal of Intelligent Manufacturing*, *25*(1), 99–107.

Yildiz, A. R., Öztürk, N., Kaya, N., & Öztürk, F. (2007). Hybrid multi-objective shape design optimization using Taguchi's method and genetic algorithm. *Structural and Multidisciplinary Optimization*, *34*(4), 317–332.

Yum, B. J., Kim, S. J., Seo, S. K., Byun, J. H., & Lee, S. H. (2013). The Taugchi robust design method: Current status and future directions. *Journal of the Korean Institute of Industrial Engineers*, *39*(5), 325–341.