Development of a new robust design methodology based on Bayesian perspectives

Nguyen Khoa Viet Truong and Sangmun Shin*

Department of Systems Management and Engineering, Inje University, Gimhae, 621-749, South Korea
E-mail: nktruong@hcmuns.edu.vn
E-mail: sshin@inje.ac.kr
*Corresponding author

Abstract: Robust design (RD), implemented in statistical and mathematical procedures to simultaneously minimise the process bias and variability, is widely used in many areas of engineering and technology to represent complex real-world industrial settings. For RD modelling and optimisation, response surface methodology (RSM) is often utilised as an estimation method to represent the functional relationship between input factors and their associated output responses. Although conventional RSM-based RD methods may offer significant advantages regarding process design, there is room for improvement. In this context, a new RD methodology is developed in this paper by integrating Bayesian principles into the RD procedure. Numerical examples and comparative studies are conducted by using two conventional RSM-based RD models and the proposed model. The results of two numerical examples demonstrate that the proposed RD method provides significantly better RD solutions in terms of the expected quality loss (EQL) than conventional methods.

Keywords: robust design; RD; response surface methodology; RSM; optimisation; expected quality loss; EQL.


Biographical notes: Nguyen Khoa Viet Truong is a Postdoctoral Researcher at Inje University. He holds a Master of Environmental Engineering from Ho Chi Minh Vietnam National University and PhD of Industrial Engineering from Inje University. Before joining PhD programme, he was a Lecturer at University of Natural Science, Ho Chi Minh city, Vietnam. His research interests are robust design, tolerance design, pharmaceutical quality by design, neural network, multi-objective optimisation and applied statistics.

Sangmun Shin is an Assistant Professor at the Department of Systems Management and Engineering and the Director of the Quality Design Laboratory at the Inje University, South Korea. He received his MS and PhD in Industrial Engineering from the Clemson University, USA. His research interests include quality engineering, robust and tolerance designs, multi-objective optimisation and pharmaceutical process design. He received his Career Development Research Award from the Korea Research Foundation.
1 Introduction

In response to the increasing imperative of global competitiveness, more effort is underway to enhance design productivity and improve quality by implementing robust design (RD) principles in quality engineering. It is also recognised that design productivity and quality improvement activities are most efficient and cost-effective when implemented during the design stage (Shin and Cho, 2009). The RD principle introduced by Taguchi (1986) includes a systematic method for applying experimental designs. RD is an enhanced process/product design methodology for reducing the variance of performance by determining the best factor settings while minimising process variability and bias (i.e., the deviation from the target value of the product). In order to execute this RD method, three major steps are often required as follows:

- **Step 1**: Experimental design to exploit information about the relationship between input factors and output responses.
- **Step 2**: Estimation of model parameters to determine the functional form of the relationship between input factors and output responses.
- **Step 3**: Optimisation to obtain robust factor settings that are less sensitive to input variation.

In an early attempt at RD research, Vining and Myers (1990) proposed a dual-response approach based on response surface methodology (RSM) as an alternative to the Taguchi approach by separately estimating the response functions of the process mean and standard deviation. Based on the dual-response principle, Del Castillo and Montgomery (1993), and Copeland and Nelson (1996) pointed out that Vining and Myers’ optimisation approach might not always guarantee optimal RD solutions, and proposed standard non-linear programming techniques, such as the generalised reduced gradient method and the Nelder-Mead simplex method that may provide better RD solutions. Regarding other aspects of response surface modelling based on estimating robust parameters and the dual-response approach, fuzzy optimisation methodologies were further developed by Khattree (1996), and Kim and Lin (1998). In terms of an optimisation framework, the dual-response model, wherein the process variability is minimised while the process mean is held at the target value, may provide a considerably large variance. Hence, the mean-squared error (MSE) model from relaxing the zero-bias assumption was proposed by Lin and Tu (1995). While allowing some process bias, the resulting process variance is less than the variance obtained from the dual-response model. Further modifications to the MSE model were discussed by Jayaram and Ibrahim (1999), Cho et al. (2000), Yue (2002), Miro-Quesada and Del Castillo (2004), and Shin and Cho (2009).

As for other extension of the MSE model, Park and Cho (2005) proposed an RD model using the weighted-least-squares method for unbalanced data. Robinson et al.

In order to perform parametric estimation by using the observed data collected from the experimental design step, the functional forms representing relationships between input factors and output responses as well as their associated model parameters have to be determined by a number of estimation methodologies, such as least squares, maximum likelihood, and Bayesian estimation. Even though RD along with RSM provides significant advantages in mathematical foundation and practicability, there is room for improvement. Based on the least-squares method, RSM requires several basic assumptions, such that errors must be independent and normally distributed with constant variance and zero expectation. If one or more basic assumptions are violated, the Gauss-Markov theorem is no longer valid. In this case, other methods, such as transformation, weighted least squares, and maximum likelihood estimation (MLE), are often considered in order to perform valid estimation. From the viewpoint of MLE, model parameters are regarded as fixed and unknown quantities, and the observed data are treated as random variables (see Martin, 2005).

Based on Bayesian estimation, the underlying assumptions are different from those of MLE and the least-squares method; thus, the unknown model parameters and observed data are treated as random variables. These assumptions may offer much flexibility in parametric estimation in many industrial situations. Under these assumptions, the quality of interest can be represented by a probability distribution of model parameters. Similar to the Bayesian approach, the posterior information represented by a probability distribution can be obtained by prior information and the relationship between the observed data and model parameters. In conventional RD approaches, both the mean and variance functions are often estimated by RSM, which utilises the least-squares method. This approach often may not satisfy the basic assumptions for validated estimation. First, in the conventional RSM-based RD approach, a variance function is estimated based on a constant error-variance assumption. Second, this approach also assumes that both the mean and variance responses belong to normal distributions although the observed variances follow Chi-square distributions (Walpole et al., 2007). To this end, inverse problem (IP) estimation methods based on the Bayesian principle are integrated with RD models in order to relax the basic assumptions of the least-squares method, by treating unknown model parameters as random variables and unobservable parameters as being drawn from a probability distribution. In order to estimate the mean function using this IP-based approach, all design points need not be homoscedastic (i.e., of the same variance). For variance function estimation, all the observations can be assumed to follow not a normal distribution but a Chi-square distribution. Also, an IP approach to RD models may provide information about the estimated parameters as a distribution. This information offers much flexibility for estimating the response functions of the process mean and dispersion.

To this end, the main objectives of this research are three-fold. Firstly, a new RD methodology including estimation, modelling, and optimisation is developed that incorporates an IP-based approach into the RD procedure in order to provide an alternative to RSM. This new IP-based RD method is far more effective than any other method for RD problems in the presence of a lack of prior information on distributions,
uncertainties in the relationship between input factors and output responses, and non-linear relationships under a large amount of data. To the best of our knowledge, this proposed approach is the first attempt in the RD literature. Secondly, a new inverse problem robust design (IPRD) model based on the MSE concept is developed in order to simultaneously minimise the process bias and variability. For illustrating the proposed method, comparative studies are conducted by using three RD models including two RSM-based RD models and the proposed IPRD model. Finally, numerical examples are examined for verification purposes. Our numerical examples clearly demonstrate that the proposed IPRD model provides significantly better solutions than conventional RSM-based RD models in terms of the expected quality loss (EQL).

2 Experimental foundation

With the aim of building quality into the design stage, RD techniques are increasingly popular in industry primarily because of their practicality. There have been many attempts to integrate Taguchi’s RD principles with well-established statistical techniques in order to model the responses directly as a function of the control factors (Vining and Myers, 1990). For obtaining data, a standard experimental format of \( n \) runs is shown Table 1. \( x, \bar{y}, s^2_{\text{obs}}, \) and \( s_{\text{obs}} \) represent the vectors of control factors, mean, variance, and standard deviation of replicated responses, respectively.

### Table 1  Experimental frame

<table>
<thead>
<tr>
<th>Runs</th>
<th>Control factor settings (X)</th>
<th>( y ) (replications)</th>
<th>( \bar{y} )</th>
<th>( s^2_{\text{obs}} )</th>
<th>( s_{\text{obs}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( y_{11}, y_{12} \ldots y_{1r} \ldots y_{1r} \bar{y}_1 )</td>
<td>( s^2_{1,\text{obs}}, s_{1,\text{obs}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( y_{21}, y_{22} \ldots y_{2r} \ldots y_{2r} \bar{y}_2 )</td>
<td>( s^2_{2,\text{obs}}, s_{2,\text{obs}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( u )</td>
<td>( y_{u1}, y_{u2} \ldots y_{ur} \ldots y_{ur} \bar{y}_u )</td>
<td>( s^2_{u,\text{obs}}, s_{u,\text{obs}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )</td>
<td>( y_{n1}, y_{n2} \ldots y_{nr} \ldots y_{nr} \bar{y}_n )</td>
<td>( s^2_{n,\text{obs}}, s_{n,\text{obs}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( y_{uv}, \bar{y}_u, s^2_{u,\text{obs}}, \) and \( s_{u,\text{obs}} \) are the \( v \)th observed response among \( r \) replications in the \( u \)th experimental run, the \( u \)th element in the vector of mean responses, the \( u \)th element in the vector of variance responses, and the \( u \)th element in the vector of standard deviation responses, respectively. Depending on the number of factors, the number of levels for each factor, and the number of experimental runs, the relationship between responses and control factors can be generalised as follows:

\[
y = f(x) = X\beta + \varepsilon
\]

where \( \beta \) and \( X \) respectively denote a vector of model parameters and an \( n \times q \) design matrix. The number of columns in the design matrix \( X \) or number of model parameters, \( q \), can be computed as:
\[ q = 1 + \sum_{j=1}^{k} C_{j+k-1}^{k-1} \]  

where \( k \) is the number of control factors.

3 Model parameters estimation

After the experimental design step, data can be used for modelling in which the model parameters of empirical functions are estimated. In this section, the conventional RSM-based method of estimation is discussed; also, a new method based on the IP approach is proposed.

3.1 Conventional RSM

For comparative purposes, conventional response surface robust design (RSRD) methods are introduced because the proposed approach plays the same role regarding RD techniques. RSM is a collection of mathematical and statistical techniques that are useful for modelling and analysing problems when the response of interest is influenced by several factors. When the exact functional relationship is not known or very complicated, the conventional least-squares method is typically used to estimate the input-response functional forms of responses in RSM (Box and Draper, 1987; Khuri and Cornell, 1987; Myers and Montgomery, 2002). Towards a comprehensive presentation of RSM, Myers (1999) provided insightful comments on the current status and future direction of RSM.

Using the output responses (i.e., \( y_{\text{obs}} \) and \( s_{\text{obs}}^2 \)), the estimated response functions of the process mean, standard deviation and variance are given by:

\[ \hat{\mu}_{\text{RSM}}(x) = \bar{x}^T \hat{\beta}_\mu \]  
\[ \hat{\sigma}_{\text{RSM}}(x) = \bar{x}^T \hat{\beta}_\sigma \]  
\[ \hat{\sigma}_{\text{RSM}}^2(x) = \bar{x}^T \hat{\beta}_{\sigma^2} \]  

where the vector \( \bar{x} \) represents a polynomial form of the input factors to explain the empirical relationships, while \( \hat{\beta}_\mu \) and \( \hat{\beta}_\sigma \) are vectors of model parameters corresponding to the output responses \( y, s_{\text{obs}} \) and \( s_{\text{obs}}^2 \). Using the conventional least-squares method, the estimated vectors (i.e., \( \hat{\beta}_\mu, \hat{\beta}_\sigma \) and \( \hat{\beta}_{\sigma^2} \)) of the model parameters \( \beta \) can be estimated as:

\[ \hat{\beta}_\mu = (X^T X)^{-1} X^T \bar{y} \]  
\[ \hat{\beta}_\sigma = (X^T X)^{-1} X^T s_{\text{obs}} \]  

and
\[ \tilde{\beta}_{\sigma^2} = (X^T X)^{-1} X^T s^2_{\text{obs}}. \]

### 3.2 The proposed IP approach

Researchers have sought to obtain knowledge from an industrial data system that includes a number of engineering problems. Based on how much information we have in advance, the system must be parameterised by discovering the minimal set of parameters that can scientifically describe the problem. Observed data can be collected in order to get specific information. For the forward modelling step, empirical functions for tentative relationships between the observed data and model parameters are built through data analysis based on a set of unknown model parameters. In order to obtain these unknown model parameters, inverse modelling is often required as a step in the modelling procedure for unknown relationships between the observed data and model parameters. In many applications, inverse modelling is used as a tool for transforming observed data to a model. Recently, IPs have arisen in many branches of science and engineering that involve statistics and mathematics where the values of some model parameters should be obtained from observed data (Tarantola, 1987, 2005; Tarantola and Valttte, 1982).

**Table 2** Comparison of RSM and IP approaches

<table>
<thead>
<tr>
<th>Compared criteria</th>
<th>RSM</th>
<th>The proposed IP approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main principle for estimation</td>
<td>• Least squares method.</td>
<td>• Bayesian perspective based on the relationship between prior and posterior information.</td>
</tr>
<tr>
<td>Classification of model parameters.</td>
<td>• Fixed values.</td>
<td>• Random variables.</td>
</tr>
<tr>
<td>Assumptions on observed data</td>
<td>• Error followed by normally independently and identically distributed with a zero mean and constant variance.</td>
<td>• Normal assumption is not required</td>
</tr>
<tr>
<td>Relationship between observed data and model parameters</td>
<td>• Deterministic point of view.</td>
<td>• Constant variances assumption is not required.</td>
</tr>
<tr>
<td>Obtained information of model parameters</td>
<td>• Deterministic values.</td>
<td>• Stochastic point of view.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Stochastic distributions.</td>
</tr>
</tbody>
</table>

RSM estimates parameters in mean and variance models by using the least-squares method based on a number of basic assumptions. The proposed IP method based on a Bayesian approach can estimate parameters without considering those assumptions. Furthermore, IP considers all the observed data and parameters as random variables whereby prior and posterior information on the data and parameters can be analysed. As a result, the outcomes of IP can provide more information than those of RSM. A brief comparison of the two approaches is shown in Table 2. One of the basic principles of RD is to seek optimal settings that can reduce the variance as well as the deviation between the mean and the target of the quality characteristic of interest among the design points. It is not reasonable to consider that all the variances of the observed data are identical. To address this problem, an IP approach can be a good candidate in which the identical-variance (i.e., homoscedasticity) assumption can be relaxed. Consequently, the
application of the IP principle to the RD optimisation procedure can overcome the disadvantages of RSM. To integrate the IP principle into the RD optimisation procedure, it is necessary to consider that all the observed data and model parameters are random variables, and that probability density functions are used to describe the information in the data space and model space. The philosophy of IP based on a Bayesian perspective can be represented by the relationship between the prior and posterior information, as shown in Figure 1.

**Figure 1** IP principle based on Bayesian perspective

![Diagram of IP principle based on Bayesian perspective](image)

4 RD optimisation

Using the results of parametric estimation, RD optimisation can then be performed in order to obtain optimal robust factor settings. In this section, two conventional RSM-based RD models and the proposed IPRD model are considered and compared in terms of the EQL criterion.

4.1 Conventional RSM-based RD method

The conventional RSM-based RD models (i.e., dual response and MSE models) by using mean function \( \hat{\mu}_{RSM} (x) \) and standard deviation function \( \hat{\sigma}_{RSM} (x) \) or variance function \( \hat{\sigma}_{RSM}^2 (x) \) can be formulated by two different optimisation frameworks.

**Model 1: Dual response model**

By separately modelling the process mean and variance using experimental data, the dual-response approach minimises \( \hat{\sigma} (x) \) subject to \( \hat{\mu} (x) = \tau \) and \( x \in \Omega \) (Vining and Myers, 1990). This model implies that first the process mean is adjusted to the target and then, the variability is minimised. The RSM-based dual-response model can be expressed as:

Minimise \( \hat{\sigma}_{RSM} (x) \)

subject to \( \hat{\mu}_{RSM} (x) = \tau, x \in \Omega \).  

(9)

In the above, \( \tau, \hat{\sigma}_{RSM} (x) \), and \( \Omega \) denote the desired target value, a function of the standard deviation, and the feasible region, respectively.
Model 2: MSE model

As Lin and Tu (1995) demonstrated, this optimisation scheme based on the zero-bias logic can be misleading due to the unrealistic constraint of forcing the estimated mean to a specific value. They proposed the MSE model to simultaneously consider the process mean and variability. In this paper, in order to conduct a comparative study based on model estimation methods using RSM and IP, a variance-based MSE model using the RSM model and incorporating the process mean and variance functions, \( \hat{\mu}_{\text{RSM}}(x) \) and \( \hat{\sigma}^2_{\text{RSM}}(x) \), can be given as follows:

\[
\begin{align*}
\text{Minimise} & \quad \left[ \hat{\mu}_{\text{RSM}}(x) - \mu \right]^2 + \hat{\sigma}^2_{\text{RSM}}(x) \\
\text{Subject to} & \quad x \in \Omega
\end{align*}
\]  

4.2 Development of IP-based RD modelling and optimisation

In an RD optimisation procedure, the principle of IP can be embedded in the model-building step wherein the mean and variance models are estimated. Conventionally, the mean and variance models in RD are built based on the assumption that the observed data at all design points are sampled from identical normal populations. Actually, this assumption is reasonable when data are sampled from normal parent distributions. However, it is not reasonable to assume that the observed variances also are of normal distributions. From this point of view, the IP approach is utilised to estimate model parameters for both the mean and variance functions assuming that the mean values belong to a normal distribution and the variance values belong to a Chi-square distribution.

The following important elements are required for integrating the IP principle in RD modelling. Firstly, all the observed data and model parameters are random variables. Secondly, their associated probability density functions are utilised to describe the information on the data and model spaces. Finally, the relationship between the prior and posterior information can be represented by the IP principle. Regarding these elements/steps, the mathematical foundations for representing the key principle of the proposed IP approach and the prior and posterior information can be described by Lemmas 1 and 2, respectively. Based on them, model parameters can be estimated statistically.

Lemma 1: Denoting \( f^{\text{d}}(d, m) \) as the probability density function of the forward problem, \( f^{\text{pri}}(d, m) \) as the prior probability density function of the observed data and model parameters, and \( f^{\text{hom}}(d, m) \) as the homogeneous density of the observed data and model parameters, using a statistical approach and Bayesian point of view, the posterior information on the observed data and model parameters can be obtained by the IP principle as follows:

\[
f^{\text{pos}}(d, m) = \text{const.} f^{\text{pri}}(d, m) f^{\text{d}}(d, m)
\]

where \( d, m, f^{\text{pos}} \) and \( f^{\text{pri}} \) denote a vector of observed data, a vector of model parameters, probability density functions describing the posterior information, and probability density functions describing the prior information, respectively. With reference to an RSM approach, the vector of observed data \( d \) can be the vector of mean responses \( y \), the vector of variance responses \( s^2 \) or the vector of standard deviations \( s \) corresponding to the vector of model parameters \( m \). A proof for this result is available in Tarantola (2005).
When the relationship between the vector of observed data \( d \) and the vector of model parameters \( m \) is linear, equations (11) and (19) can have a simpler form as shown in Lemma 2.

**Lemma 2:** Assuming that the observed data have no prior information and their populations are independent of those of the model parameters, a specific form of equations (11) and (19) introduced in Lemma 1 is applicable. Also, assuming that the empirical relationship has a linear form, viz.,

\[
\mathbf{d} = \mathbf{Xm}
\]  

Equations (11) and (19) can be rewritten as:

\[
f^{Post}(\mathbf{m}) = \psi f^{pri}(\mathbf{m}) f^{pri}(\mathbf{Xm})
\]

where the normalised constant \( \psi \) is defined as:

\[
\psi = \left( \int f^{pri}(\mathbf{m}) f^{pri}(\mathbf{Xm}) d\mathbf{m} \right)^{-1}.
\]

In general, the expectation vector of the model parameters and covariance matrix of the model parameters representing posterior information on the model parameters is calculated by integrating the posterior probability density function as follows:

\[
\hat{\mathbf{m}} = E(\mathbf{m}) = \int \mathbf{m} f^{Post}(\mathbf{m}) d\mathbf{m}
= \psi \int \mathbf{m} f^{pri}(\mathbf{m}) f^{pri}(\mathbf{Xm}) d\mathbf{m}
\]

and

\[
\hat{C}_M = \int (\mathbf{m} - E(\mathbf{m}))(\mathbf{m} - E(\mathbf{m}))^T f^{Post}(\mathbf{m}) d\mathbf{m}
= \psi \int (\mathbf{m} - E(\mathbf{m}))(\mathbf{m} - E(\mathbf{m}))^T f^{pri}(\mathbf{m}) f^{pri}(\mathbf{Xm}) d\mathbf{m}.
\]

In the above, \( \hat{\mathbf{m}} \) and \( \hat{C}_M \) respectively denote the vector of expectations and covariance matrix of the model parameters. The proof for this result is also available in Tarantola (2005).

### 4.2.1 Development of a mean model on IPRD

Based on Lemmas 1 and 2, a new mean model for the IPRD procedure is developed by estimating the associated parameters. The key concepts are addressed in Lemma 3 and Propositions A.1 through A.4.

**Lemma 3:** The results of Lemmas 1 and 2 can be utilised for both the mean and variance models. Assuming that an empirical relationship between the control factors and their associated mean responses has a generalised linear form and substituting \( \mathbf{y} \) for the vector \( \mathbf{d} \) in equations (11) and (19), and (20) and (12), the estimated response function can be identified as:

\[
\mathbf{y} = \mathbf{Xm}
\]
Development of a new robust design methodology

where \( \mathbf{m}_m \) is the vector of model parameters for the mean function. When the observed responses are of normal distributions, their mean values at the design points are also of normal distributions with mean vector \( \mathbf{y} \) and covariance matrix \( \mathbf{C}_D \). If there is no prior information on the model parameters, \( f^{\alpha}(\mathbf{m}) \) in equations (14) and (22) will be a constant, and the posterior information of the model parameters can then be derived as:

\[
f^{\text{Pos}}(\mathbf{m}_m) = \psi \exp\left[-\frac{1}{2}(\mathbf{Xm}_m - \bar{\mathbf{y}})^T \mathbf{C}_D^{-1} (\mathbf{Xm}_m - \bar{\mathbf{y}})\right],
\]

(17)

where \( \psi = \{2\pi^n \det(\mathbf{C}_D)^{-1/2} \} \) is a normalised constant.

Based on Lemma 2, in the case of mean function responses, the expectation and covariance matrix of the model parameters \( \mathbf{m}_m \) for the mean function can be estimated by using the posterior probability density function as detailed in Propositions A1, A2, A3 and A4.

These propositions and their associated proofs which discussed in Appendix A identify mathematical foundations for obtaining the expectations and covariances of the model parameters.

**Proposition A.1:** With the assumptions discussed in Lemma 3, the posterior information of the model parameters can be obtained as follows:

\[
\bar{\mathbf{m}}_m = \left(\mathbf{X}^T \mathbf{C}_D^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \mathbf{C}_D^{-1} \bar{\mathbf{y}}\right)
\]

(18)

\[
\bar{\mathbf{C}}_M = \left(\mathbf{X}^T \mathbf{C}_D^{-1} \mathbf{X}\right)^{-1}.
\]

(19)

**Proposition A.2:** In equations (17) and (25), the inside matrix of the exponential function can be expressed as

\[
(\mathbf{Xm}_m - \bar{\mathbf{y}})^T \mathbf{C}_D^{-1} (\mathbf{Xm}_m - \bar{\mathbf{y}}) = (\mathbf{m}_m - \bar{\mathbf{m}}_m)^T \bar{\mathbf{C}}_M (\mathbf{m}_m - \bar{\mathbf{m}}_m)
\]

(20)

where

\[
\bar{\mathbf{C}}_M = \left(\mathbf{X}^T \mathbf{C}_D^{-1} \mathbf{X}\right)^{-1} \text{ and } \bar{\mathbf{m}}_m = \left(\mathbf{X}^T \mathbf{C}_D^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \mathbf{C}_D^{-1} \bar{\mathbf{y}}\right).
\]

**Proposition A.3:** Denoting the posterior information of the model space in a joint-distribution form, which is

\[
f^{\text{Pos}}(\mathbf{m}_m) = \psi \exp\left[-\frac{1}{2}(\mathbf{m}_m - \bar{\mathbf{m}}_m)^T \bar{\mathbf{C}}_M (\mathbf{m}_m - \bar{\mathbf{m}}_m)\right],
\]

(21)

the mean vector of the model parameters representing the posterior information of the model space is defined by the observed mean vector of the model parameters, i.e.,

\[
\bar{\mathbf{m}}_m = \bar{\mathbf{m}}_m.
\]

(22)

**Proposition A.4:** We can denote the posterior information of the model parameters as a joint-distribution form, which is

\[
f^{\text{Pos}}(\mathbf{m}_m) = \psi \cdot \exp\left[-\frac{1}{2}(\mathbf{m}_m - \bar{\mathbf{m}}_m)^T \bar{\mathbf{C}}_M (\mathbf{m}_m - \bar{\mathbf{m}}_m)\right]
\]

(23)
where the covariance matrix of the model parameters is defined as
\[ \hat{C}_M = \tilde{C}_M. \] (24)

4.2.2 Development of a variance model on IPRD

Similar to the proposed mean model development procedure identified in Lemmas 1 and 2, the posterior information of the model parameters for the variance response function can be estimated by the following propositions (Propositions B.1 and B.2).

Proposition B.1: We can denote the posterior information of the model parameters as a joint-distribution form, which is
\[
\hat{f}^{\text{pos}}(\mathbf{m}_v) = \frac{1}{\left(2^{-r/2} \Gamma\left(r-1/2\right)\right)^{1/2}} \prod_{u=1}^{n} \left(\frac{(r-1)s_u^2}{s_{u,\text{obs}}^2}\right)^{\frac{r-1}{2}} \cdot \exp \left[ -\sum_{u=1}^{n} \frac{(r-1)s_u^2}{2s_{u,\text{obs}}^2} \right].
\] (25)

Proposition B.2: Using the assumptions discussed in Proposition B.1, the posterior information of the model parameters can be obtained as
\[
\hat{m}_v = \left(K^T K\right)^{-1} K^T F
\] (26)
where
\[
K = \left[\begin{array}{c}
(r-1)(X_{11})^2 \\
(r-1)(X_{21})^2 \\
\vdots \\
(r-1)(X_{n1})^2 \\
\frac{s_{1,\text{obs}}^2}{s_{1,\text{obs}}^2} \\
\frac{s_{2,\text{obs}}^2}{s_{2,\text{obs}}^2} \\
\vdots \\
\frac{s_{n,\text{obs}}^2}{s_{n,\text{obs}}^2}
\end{array}\right]^T
\]
is a modified form of the design matrix \(X\). Details can be seen further as well as Propositions B.1 and B.2 in Appendix B.

4.2.3 The proposed IPRD model

The customisation of the general IP framework to RD modelling and optimisation can be performed by estimating the response functions of the process mean and variance similar to the RSM procedure as shown in equations (3) and (4). For IPRD modelling, the estimated response functions of the process mean and variance can be obtained as
\[
\hat{\mu}_{IP}(x) = \tilde{x}^T \hat{m}_m
\] (27)
\[
\hat{\sigma}_{IP}^2(x) = \tilde{x}^T \hat{m}_v
\] (28)
where the vector \( \tilde{x} \) represents a polynomial form of the input factors to explain their empirical relationships. For example, assuming that a second order is chosen with three control factors, the vector \( \tilde{x} \) can be defined as
\[
\tilde{x}^T = [1 \ x_1 \ x_2 \ x_3 \ x_1^2 \ x_2^2 \ x_3^2 \ x_1x_2 \ x_1x_3 \ x_2x_3].
\] (29)
Model 3: The proposed IPRD model

Using equations (27) and (28), a generalised IPRD model using the response functions of the process mean and variance based on the MSE concept can be formulated as follows:

Minimise \[ \left[ \hat{\mu}_P(x) - r \right]^2 + \sigma_{IP}^2(x) \]

Subject to \[ x \in \Omega \]  

where \[
\hat{\mu}_P(x) = \bar{x}^T \left( X^T C^{-1} X \right)^{-1} X^T C^{-1} y
\]
\[
\hat{\sigma}_{IP}^2(x) = \bar{x}^T \left( K^T K \right)^{-1} K^T F
\]

Figure 2 An overview of the proposed IPRD procedure (see online version for colours)
Based on the general RD procedure including three sequential steps (i.e., experimental design, model parameter estimation, and RD optimisation), an overview of the proposed IPRD is illustrated in Figure 2.

4.3 Model comparison and comparative criterion

In order to compare the three RD models including the two conventional RD models (i.e., the dual-response and MSE models) and the proposed IPRD model, the model schemes and properties for each model are comparatively discussed in Table 3. As shown in Table 3, the two conventional RD models have the same mean and variance functions based on RSM but different optimisation schemes. Although the proposed IPRD model has the same optimisation scheme, the estimated mean and variance functions based on the proposed IP approach are different from those of the conventional models.

Table 3 Models comparison and discussions

<table>
<thead>
<tr>
<th>Models</th>
<th>Names</th>
<th>Optimisation models</th>
<th>Model schemes and discussions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DR model</td>
<td>Min $\hat{\sigma}<em>{RSM}(x)$ s.t $\hat{\mu}</em>{RSM}(x) = \tau, x \in \Omega$</td>
<td>• Optimisation model based on a dual response model estimation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>where $\hat{\mu}_{RSM}(x) = \hat{x}^T (X^T X)^{-1} X^T \bar{y}$</td>
<td>• Mean model $\hat{\mu}_{RSM}(x)$ is estimated from $\bar{y}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\sigma}<em>{RSM}(x) = \hat{x}^T (X^T X)^{-1} X^T s</em>{obs}$</td>
<td>• Variance model $\hat{\sigma}_{RSM}(x)$ is estimated from $s$</td>
</tr>
<tr>
<td>2</td>
<td>MSE model</td>
<td>Min $[\hat{\mu}<em>{RSM}(x) - \tau]^2 + \hat{\sigma}</em>{RSM}^2(x)$ s.t $x \in \Omega$</td>
<td>• Optimisation model based on a MSE model estimation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>where $\hat{\mu}_{RSM}(x) = \hat{x}^T (X^T X)^{-1} X^T \bar{y}$</td>
<td>• Mean model $\hat{\mu}_{RSM}(x)$ is estimated from $\bar{y}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\sigma}<em>{RSM}^2(x) = \hat{x}^T (X^T X)^{-1} X^T s</em>{obs}^2$</td>
<td>• Variance model is estimated from $s$</td>
</tr>
<tr>
<td>3</td>
<td>The proposed IPRD model</td>
<td>Min $[\hat{\mu}<em>{IP}(x) - \tau]^2 + \hat{\sigma}</em>{IP}^2(x)$ s.t $x \in \Omega$</td>
<td>• Optimisation model based on an inverse problem-based estimation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>where $\hat{\mu}_{IP}(x) = \hat{x}^T (X^T C_B^{-1} X)^{-1} X^T C_B^{-1} \bar{y}$</td>
<td>• Mean model $\hat{\mu}_{IP}(x)$ is estimated from $\bar{y}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\sigma}_{IP}^2(x) = \hat{x}^T (K^T K)^{-1} K^T F$</td>
<td>• $C_B$ has different variance at design points are estimated by variance of design points</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• When $C_B = const.I$, mean model $\hat{\mu}_{IP}(x)$ will become mean models 1 and 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Variance model $\hat{\sigma}_{IP}^2(x)$ is estimated from the matrix $K$ that contains variance information of design points</td>
</tr>
</tbody>
</table>
To perform a comparative study, a number of criteria for robust solutions are discussed in the related literature. Among these criteria, the EQL is the most commonly used comparative criterion. One advantage of using this criterion is that a loss can be decomposed by the process mean and dispersion (i.e., variance and standard deviation). If the loss function \( L(y) \) is other than quadratic, then higher-order moments of the random variable \( Y \) are often necessary in order to evaluate the expected loss. More detailed discussions about this loss function under different environments can be found in Box et al. (1988) and Spring (1991).

Depending on the forms of the process mean and variance functions, the real order of the EQL function can be fourth or higher. If we let \( L(y) \) be a measure of losses associated with the quality characteristic \( Y \), whose target value is \( \tau \), then the quadratic loss function is given by

\[
L(y) = \theta (y - \tau)^2
\]

where \( \theta \) is a positive loss coefficient. Then, EQL is given by

\[
E[L(y)] = \theta \left[ (\mu - \tau)^2 + \sigma^2 \right]
\]

(31)

where \( \mu \) and \( \sigma^2 \) denote the process mean and variance, respectively. Using EQL, we compare the optimal solutions of the three models.

5 Numerical examples

5.1 The printing study

One of the most well-known examples in the field of RD is the printing study used by Box and Draper (1987), Vining and Myers (1990), Lin and Tu (1995), Copeland and Nelson (1996), and Cho et al. (2000). This numerical example is also selected in this paper for comparative purposes. The printing study was conducted to investigate the effects of the speed (\( x_1 \)), pressure (\( x_2 \)), and distance (\( x_3 \)) on the ability of a printing machine to apply coloured inks to package levels (\( y \)). There are 27 runs of a three-level factorial design of three control factors with three replications at each design point used to fit the responses. Table 4 displays the experimental data for the printing study.

5.1.1 Results of estimation

The results of the comparative study including the values of the model parameters for four response functions can be seen in Table 5. This table provides the vectors of the model parameters of two response functions (\( \hat{\mu}_{RSM}(x) \) and \( \hat{\sigma}_{RSM} \)) estimated by using RSM and the other two response functions of the process mean and variance (\( \hat{\mu}_{IP}(x) \) and \( \hat{\sigma}_{IP}^5(x) \)) estimated by using IP.

Using RSM, the vectors of the model parameters for the fitted response functions of the process mean and variance for Models 1 and 2 are given by equations (3) and (4). Using the proposed IP approach, the vectors of model parameters for the fitted response functions of the process mean and variance for Model 3 are given by equations (18) and (26).
Table 4  The printing study data

<table>
<thead>
<tr>
<th>Runs</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$y$</th>
<th>$y^2$</th>
<th>$s$</th>
<th>$s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>34</td>
<td>10</td>
<td>28</td>
<td>24,000</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>–1</td>
<td>–1</td>
<td>115</td>
<td>116</td>
<td>130</td>
<td>120,333</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>–1</td>
<td>–1</td>
<td>192</td>
<td>186</td>
<td>263</td>
<td>213,667</td>
</tr>
<tr>
<td>4</td>
<td>–1</td>
<td>0</td>
<td>–1</td>
<td>82</td>
<td>88</td>
<td>88</td>
<td>86,000</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>–1</td>
<td>44</td>
<td>178</td>
<td>188</td>
<td>136,667</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>–1</td>
<td>322</td>
<td>350</td>
<td>350</td>
<td>340,667</td>
</tr>
<tr>
<td>7</td>
<td>–1</td>
<td>1</td>
<td>–1</td>
<td>141</td>
<td>110</td>
<td>86</td>
<td>112,333</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>–1</td>
<td>259</td>
<td>251</td>
<td>259</td>
<td>256,333</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>–1</td>
<td>290</td>
<td>280</td>
<td>245</td>
<td>271,667</td>
</tr>
<tr>
<td>10</td>
<td>–1</td>
<td>–1</td>
<td>0</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81,000</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>–1</td>
<td>0</td>
<td>90</td>
<td>122</td>
<td>93</td>
<td>101,667</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>–1</td>
<td>0</td>
<td>319</td>
<td>376</td>
<td>376</td>
<td>357,000</td>
</tr>
<tr>
<td>13</td>
<td>–1</td>
<td>0</td>
<td>0</td>
<td>180</td>
<td>180</td>
<td>154</td>
<td>171,333</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>372</td>
<td>372</td>
<td>372</td>
<td>372,000</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>541</td>
<td>568</td>
<td>396</td>
<td>501,667</td>
</tr>
<tr>
<td>16</td>
<td>–1</td>
<td>1</td>
<td>0</td>
<td>288</td>
<td>192</td>
<td>312</td>
<td>264,000</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>432</td>
<td>336</td>
<td>513</td>
<td>427,000</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>713</td>
<td>725</td>
<td>754</td>
<td>730,667</td>
</tr>
<tr>
<td>19</td>
<td>–1</td>
<td>–1</td>
<td>1</td>
<td>364</td>
<td>99</td>
<td>199</td>
<td>220,667</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>–1</td>
<td>1</td>
<td>232</td>
<td>221</td>
<td>266</td>
<td>239,667</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>–1</td>
<td>1</td>
<td>408</td>
<td>415</td>
<td>443</td>
<td>422,000</td>
</tr>
<tr>
<td>22</td>
<td>–1</td>
<td>0</td>
<td>1</td>
<td>182</td>
<td>233</td>
<td>182</td>
<td>199,000</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>507</td>
<td>515</td>
<td>434</td>
<td>485,333</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>846</td>
<td>535</td>
<td>640</td>
<td>673,667</td>
</tr>
<tr>
<td>25</td>
<td>–1</td>
<td>1</td>
<td>1</td>
<td>236</td>
<td>126</td>
<td>168</td>
<td>176,667</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>660</td>
<td>440</td>
<td>403</td>
<td>501,000</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>878</td>
<td>991</td>
<td>1161</td>
<td>1,010,000</td>
</tr>
</tbody>
</table>

Table 5  Coefficients of the estimated response functions

<table>
<thead>
<tr>
<th>Treatment combinations</th>
<th>$\hat{\beta}_\mu$</th>
<th>$\hat{\beta}_\alpha$</th>
<th>$\hat{\beta}_m$</th>
<th>$\hat{\beta}_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>327.630</td>
<td>34.883</td>
<td>372</td>
<td>0.0001</td>
</tr>
<tr>
<td>$x_1$</td>
<td>177.000</td>
<td>11.527</td>
<td>176,077</td>
<td>20.228</td>
</tr>
<tr>
<td>$x_2$</td>
<td>109.426</td>
<td>15.323</td>
<td>130,1378</td>
<td>279.868</td>
</tr>
<tr>
<td>$x_3$</td>
<td>131.463</td>
<td>29.190</td>
<td>120,486</td>
<td>360.358</td>
</tr>
<tr>
<td>$x_1^2$</td>
<td>32.000</td>
<td>4.204</td>
<td>–7.010</td>
<td>270.353</td>
</tr>
<tr>
<td>$x_2^2$</td>
<td>–22.389</td>
<td>–1.316</td>
<td>–24.738</td>
<td>203.989</td>
</tr>
<tr>
<td>$x_3^2$</td>
<td>–29.056</td>
<td>16.778</td>
<td>–44.373</td>
<td>212.803</td>
</tr>
<tr>
<td>$x_1 x_2$</td>
<td>66.028</td>
<td>7.719</td>
<td>46.962</td>
<td>–174.246</td>
</tr>
<tr>
<td>$x_1 x_3$</td>
<td>75.472</td>
<td>5.109</td>
<td>63.182</td>
<td>–90.526</td>
</tr>
<tr>
<td>$x_2 x_3$</td>
<td>43.583</td>
<td>14.082</td>
<td>59.835</td>
<td>314.287</td>
</tr>
</tbody>
</table>
5.1.2 Results of comparative studies

Using the target of 500 and EQL as the comparative criterion, Table 6 comparatively shows the optimisation results, such as the optimal factor settings, mean, bias, variance, standard deviation, and EQL.

Table 6  Comparative study results for printing study

<table>
<thead>
<tr>
<th>Models</th>
<th>Model name</th>
<th>Optimal settings</th>
<th>Mean</th>
<th>Bias</th>
<th>STD</th>
<th>VAR</th>
<th>EQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DR</td>
<td>1.000 0.116 -0.258</td>
<td>500.00</td>
<td>0.00</td>
<td>45.11</td>
<td>2,034.79</td>
<td>2,034.79</td>
</tr>
<tr>
<td>2</td>
<td>MSE</td>
<td>1.000 0.072 -0.250</td>
<td>494.67</td>
<td>5.33</td>
<td>44.47</td>
<td>1,977.54</td>
<td>2,005.92</td>
</tr>
<tr>
<td>3</td>
<td>The proposed IPRD</td>
<td>0.618 0.287 -0.124</td>
<td>499.09</td>
<td>0.91</td>
<td>11.66</td>
<td>136.03</td>
<td>136.86</td>
</tr>
</tbody>
</table>

As shown in Table 6, the results of the RD model provide zero bias with considerably large variance $\sigma^2(x) = 2,034.79$ because the process mean is fixed at the target value. In the results of the MSE model, the process variance $\hat{\sigma}^2(x) = 1,977.54$ is smaller than that of the DR model owing to a bias of $|\hat{\mu}(x) - \tau| = 5.33$. The results of the proposed IPRD model yield a significantly smaller variance of $\sigma^2(x) = 136.03$ compared with the two conventional RD models (i.e., the DR and MSE models). Based on EQL, the proposed IPRD model is much better than the two conventional RD models in this example: EQL in the DR model = 2,34.79, 2,005.92 in the MSE model, and 136.86 in the proposed
IPRD model. In order to demonstrate the optimisation results in terms of criterion spaces, the process mean vs. variance and the squared bias vs. variance of the three models are illustrated in Figures 3, 4, and 5.

Figure 4  Criterion space for Model 2 (see online version for colours)

Figure 5  Criterion space for Model 3 (see online version for colours)
5.2 The Catapult study

A Roman-style catapult experiment was introduced by Luner (1994) using a dual-response approach in estimation and by Kim and Lin (1998) using a fuzzy modelling approach based on dual-response optimisation. In this example, three variables were used, viz., $x_1$ (arm length), $x_2$ (stop angle), and $x_3$ (pivot height), to predict the distance to the point where a projectile landed from the base of the catapult. A central composite design with three factors and three replications is used in this experiment. The data for the catapult study are given in Table 7.

Table 7  Catapult study data

<table>
<thead>
<tr>
<th>Runs</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$y$</th>
<th>$\bar{y}$</th>
<th>$s$</th>
<th>$s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>39</td>
<td>34</td>
<td>42</td>
<td>38.333</td>
</tr>
<tr>
<td>2</td>
<td>–1</td>
<td>1</td>
<td>–1</td>
<td>80</td>
<td>71</td>
<td>91</td>
<td>80.667</td>
</tr>
<tr>
<td>3</td>
<td>–1</td>
<td>1</td>
<td>1</td>
<td>52</td>
<td>44</td>
<td>45</td>
<td>47.000</td>
</tr>
<tr>
<td>4</td>
<td>–1</td>
<td>1</td>
<td>1</td>
<td>97</td>
<td>68</td>
<td>60</td>
<td>75.000</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>–1</td>
<td>–1</td>
<td>60</td>
<td>53</td>
<td>68</td>
<td>60.333</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>113</td>
<td>104</td>
<td>127</td>
<td>114.667</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>–1</td>
<td>78</td>
<td>64</td>
<td>65</td>
<td>69.000</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>130</td>
<td>79</td>
<td>75</td>
<td>94.667</td>
</tr>
<tr>
<td>9</td>
<td>–1.682</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>51</td>
<td>60</td>
<td>56.667</td>
</tr>
<tr>
<td>10</td>
<td>1.682</td>
<td>0</td>
<td>0</td>
<td>115</td>
<td>102</td>
<td>117</td>
<td>111.333</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>–1.682</td>
<td>0</td>
<td>50</td>
<td>43</td>
<td>57</td>
<td>50.000</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1.682</td>
<td>0</td>
<td>88</td>
<td>49</td>
<td>43</td>
<td>60.000</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>–1.682</td>
<td>54</td>
<td>50</td>
<td>60</td>
<td>54.667</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1.682</td>
<td>122</td>
<td>109</td>
<td>119</td>
<td>116.670</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87</td>
<td>78</td>
<td>89</td>
<td>84.667</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>86</td>
<td>79</td>
<td>85</td>
<td>83.333</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>81</td>
<td>87</td>
<td>85.333</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>89</td>
<td>82</td>
<td>87</td>
<td>86.000</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>86</td>
<td>79</td>
<td>88</td>
<td>84.333</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>79</td>
<td>90</td>
<td>85.667</td>
</tr>
</tbody>
</table>

Table 8  Comparative study results for Catapult study

<table>
<thead>
<tr>
<th>Models</th>
<th>Model name</th>
<th>Optimal settings</th>
<th>Mean $\hat{\mu}(x)$</th>
<th>Bias $\hat{\mu}(x) - \bar{y}$</th>
<th>STD $\hat{\sigma}(x)$</th>
<th>$VAR \hat{\sigma}^2(x)$</th>
<th>EQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DR</td>
<td>0.229 0.747 –0.651</td>
<td>100.00</td>
<td>0.00</td>
<td>5.66</td>
<td>32.00</td>
<td>32.00</td>
</tr>
<tr>
<td>2</td>
<td>MSE</td>
<td>0.225 0.710 –0.638</td>
<td>99.25</td>
<td>0.75</td>
<td>5.55</td>
<td>30.85</td>
<td>31.42</td>
</tr>
<tr>
<td>3</td>
<td>Proposed IP</td>
<td>0.287 0.514 –0.044</td>
<td>99.59</td>
<td>0.41</td>
<td>5.22</td>
<td>27.26</td>
<td>27.43</td>
</tr>
</tbody>
</table>
Figure 6  Criterion space for Model 1 (see online version for colours)

Figure 7  Criterion space for Model 2 (see online version for colours)
Development of a new robust design methodology

Figure 8  Criterion space for Model 3 (see online version for colours)

Following the estimation of the vectors of model parameters through RSM and the proposed IPRD approach, the optimisation results of the three RD models for this catapult experiment are summarised in Table 8.

In this case study, the target value is identified as 100. Based on the EQL criterion, the results of this experimental study also indicate that the proposed IPRD model provides slightly better solutions compared with the DR and MSE models: 32.00 for the RD model, 31.42 for the MSE model, and 27.43 for the proposed IPRD model. In addition, Figures 6, 7, and 8 illustrate the criterion spaces of the three models.

6 Conclusions

In this paper, by using the concept of MSE to relax the basic assumptions for the least-squares method, the customisation of IP estimation methods to RD problems based on a Bayesian point of view was addressed. Then, IPRD models were developed in order to simultaneously minimise the process bias and variability. Using the three RD models including the two conventional RSM-based RD models and the proposed IPRD model, comparative studies were conducted. Based on numerical examples, the proposed IPRD models were shown to provide significantly better solutions than the conventional RD models in terms of the EQL criterion. The comparative studies clearly showed how to apply the proposed IP estimation method in the RD procedure. The proposed IPRD model provides the estimated parameters as distributions. This information has ample flexibility for estimating the process mean and variance functions. Based on the results of the comparative analysis, the proposed IPRD models also provided significantly low variance in the particular examples by applying different variances at all design points.
These significant results imply that the proposed methods provide more robust solutions than conventional RSM approaches because IP-based RD methods can exploit much more information on datasets by using probabilistic and distributional approaches. Further extensions of the customised IP approach to RD problems can be realised by considering high-order response functions and noise factors.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (20100104).

References


Appendix A

Model parameters of mean response model

Proposition A.1: With the assumptions discussed in Lemma 3, the posterior information of the model parameters can be obtained as follows:

\[
\begin{align*}
\tilde{m}_m &= (X^T C_D^{-1} X)^{-1} (X^T C_D^{-1} \tau) \\
\tilde{C}_M &= (X^T C_D^{-1} X)^{-1}.
\end{align*}
\]

Proof: According to equations (14) and (22), and (17) and (25), the posterior information of the model parameters can be estimated as follows.
\[ \hat{m}_m = E(m_m) = \int m_m f_{\text{pos}}(m_m) \, dm_m \]
\[ = \psi \int m_m \exp \left[ -\frac{1}{2} (Xm_m - \bar{y})^T C_D^{-1} (Xm_m - \bar{y}) \right] \, dm_m \]  
(A.3)

and

\[ \hat{C}_M = \int (m_m - E(m_m)) (m_m - E(m_m))^T f_{\text{pos}}(m_m) \, dm_m \]
\[ = \psi \int (m_m - E(m_m)) (m_m - E(m_m))^T \]
\[ \exp \left[ -\frac{1}{2} (Xm_m - \bar{y})^T C_D^{-1} (Xm_m - \bar{y}) \right] \, dm_m. \]  
(A4)

By applying the results of Proposition A.2, the posterior information of the model parameters as addressed in equations (A.3) and (A.4) can be re-identified as

\[ \hat{m}_m = \psi \int m_m \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T C_M^{-1} (m_m - \bar{m}_m) \right] \, dm_m \]  
(A.5)

and

\[ \hat{C}_M = \psi \int (m_m - E(m_m)) (m_m - E(m_m))^T \]
\[ \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T C_M^{-1} (m_m - \bar{m}_m) \right] \, dm_m. \]  
(A6)

where \( C_M = (X^T C_D^{-1} X)^{-1} \) and \( \bar{m}_m = (X^T C_D^{-1} X)^{-1} (X^T C_D^{-1} \bar{y}) \). By substituting the results of Propositions A.3 and A.4, the posterior information of the model parameters can then be estimated by the following equations, viz., (A.7) and (A.8).

\[ \hat{m}_m = (X^T C_D^{-1} X)^{-1} (X^T C_D^{-1} \bar{y}) \]  
(A.7)

\[ \hat{C}_M = (X^T C_D^{-1} X)^{-1}. \]  
(A.8)

**Proposition A.2:** In equations (17) and (25), the inside matrix of the exponential function can be expressed as

\[ (Xm_m - \bar{y})^T C_D^{-1} (Xm_m - \bar{y}) = (m_m - \bar{m}_m)^T C_M^{-1} (m_m - \bar{m}_m) \]  
(A.9)

where

\[ C_M = (X^T C_D^{-1} X)^{-1} \]  
and \( \bar{m}_m = (X^T C_D^{-1} X)^{-1} (X^T C_D^{-1} \bar{y}) \).

**Proof:** It is clear that \( (X^T C_D^{-1} X)^{-1} \) \( X^T C_D^{-1} X = I \). The left-hand side of equation (A.9) can be extended as follows:
Development of a new robust design methodology

(A.10)

By denoting $\bar{C}_M = (X^T C_D^{-1} X)^{-1}$ and $\bar{m}_m = (X^T C_D^{-1} X)^{-1} (X^T C_D^{-1} Y)$, equation (A.10) can be expressed as:

\[
(X_m - \bar{y})^T C_D^{-1} (X_m - \bar{y}) = \left[ m_m - (X^T C_D^{-1} X)^{-1} X^T C_D^{-1} X \right] \left[ m_m - (X^T C_D^{-1} X)^{-1} X^T C_D^{-1} Y \right]
\]

(A.11)

Proposition A.3: Denoting the posterior information of the model space in a joint-distribution form, which is

\[
f^{\text{Pos}} (m_m) = \psi \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \bar{C}_M^{-1} (m_m - \bar{m}_m) \right],
\]

(A.12)

the mean vector of the model parameters representing the posterior information of the model space is defined by the observed mean vector of the model parameters, i.e.,

\[
\tilde{m}_m = m_m.
\]

(A.13)

Proof: The mean vector of the model parameters $\tilde{m}_m$ can be estimated by taking the expectation of $m_m$ as follows:

\[
\tilde{m}_m = E(m_m) = \int m_m f^{\text{Pos}} (m_m)
\]

(A.14)

where

\[
f^{\text{Pos}} (m_m) = \psi \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \bar{C}_M^{-1} (m_m - \bar{m}_m) \right]
\]

and where $m_m f^{\text{Pos}} (m_m)$ in the right-hand side of equation (A.14) can be transformed as follows:

\[
m_m f^{\text{Pos}} (m_m) = \psi m_m \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \bar{C}_M^{-1} (m_m - \bar{m}_m) \right]
\]

(A.15)

Based on equation (A.15), equation (A.14) can be rewritten as:
\[ \tilde{m}_m = E(m_m) = \psi \int (m_m - \bar{m}_m) \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] \]

(A.16)

The first part of the right-hand side in equation (A.16) can be identified as:

\[ \psi \int (m_m - \bar{m}_m) \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] = \psi \int \tilde{m} \exp \left[ -\frac{1}{2} \tilde{m}^T \tilde{C}_M^{-1} \tilde{m} \right] \]

(A.17)

where \( m^* = (m_m - \bar{m}_m) \). In addition, it is also the case that

\[ \psi \int m^* \exp \left[ -\frac{1}{2} m^* \tilde{C}_M^{-1} m^* \right] = 0, \]

(A.18)

because \( \psi \int \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] = 1 \) and \( \bar{m}_m = \text{const} \). Thus, equation (A.16) can be simplified as:

\[ \tilde{m}_m = E(m_m) = \bar{m}_m. \]

(A.19)

**Proposition A.4:** We can denote the posterior information of the model parameters as a joint-distribution form, which is

\[ f^{\text{Post}}(m_m) = \psi \cdot \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] \]

(A.20)

where the covariance matrix of the model parameters is defined as

\[ \tilde{C}_M = \tilde{C}_M. \]

(A.21)

**Proof:** From equation (A.6), the covariance matrix of the model parameters, \( \tilde{C}_M \), can be estimated as:

\[ \tilde{C}_M = \psi \int (m_m - E(m_m))(m_m - E(m_m))^T \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] / m_m \]

(A.22)

Using \( E(m_m) = \bar{m}_m \) in equation (A.19), equation (A.22) can be re-identified as follows:

\[ \tilde{C}_M = \psi \int (m_m - \bar{m}_m)(m_m - \bar{m}_m)^T \exp \left[ -\frac{1}{2} (m_m - \bar{m}_m)^T \tilde{C}_M^{-1} (m_m - \bar{m}_m) \right] / m_m \]

(A.23)
Letting \( \mathbf{m}^* = (\mathbf{m}_m - \bar{\mathbf{m}}_m) \) similar to Proposition A.3, equation (A.23) can be simplified as

\[
\hat{\mathbf{C}}_M = \psi \int \mathbf{m}^* \mathbf{m}^* \exp \left[ -\frac{1}{2} \mathbf{m}^* \mathbf{C}^{-1}_m \mathbf{m}^* \right] d\mathbf{m}_m. \tag{A.24}
\]

Since equation (A.24) can be used to estimate the second moment of \( \mathbf{m}^* \), the covariance matrix can be identified as

\[
\hat{\mathbf{C}}_M = E(\mathbf{m}^* \mathbf{m}^*) = \mathbf{C}_M. \tag{A.25}
\]

Once the vector of model parameters can be estimated based on the results of Proposition A.1, the mean model can be proposed as in equation (A.26):

\[
\hat{\mu}_{IP}(\mathbf{x}) = \mathbf{x}^T \hat{\mathbf{m}}_m, \tag{A.26}
\]

where the vector of model parameters \( \hat{\mathbf{m}}_m \) can be estimated by equation (A.1) as follows:

\[
\hat{\mathbf{m}}_m = (X^T \mathbf{C}_D^{-1} X)^{-1} X^T \mathbf{C}_D^{-1} \mathbf{y}.
\]

**Appendix B**

**Inverse problem for variance response model**

**Proposition B.1:** We can denote the posterior information of the model parameters as a joint-distribution form, which is

\[
f^{pos}(\mathbf{m}) = \frac{1}{2^{r/2} \Gamma\left(\frac{r-1}{2}\right)} \prod_{u=1}^{\eta} \left( \frac{(r-1)s^2_u}{s^2_{u\_obs}} \right)^{r_u-1} \exp \left[ -\sum_{u=1}^{\eta} \frac{(r-1)s^2_u}{2s^2_{u\_obs}} \right]. \tag{B.1}
\]

**Proof:** By substituting \( \mathbf{s}^2 \) for the vector \( \mathbf{d} \) in equations (11) and (19), and (12) and (20), the empirical relationship between the variance responses and control factors can be formulated in a linear form as follows:

\[
\mathbf{s}^2 = \mathbf{Xm}_v, \tag{B.2}
\]

where \( \mathbf{m}_v \) denotes the vector of model parameters of the variance function. When the observed responses are of a normal distribution, the variance values at all design points will belong to a Chi-square distribution with a mean vector of \( \mathbf{s}^2_{obs} \) (Walpole et al., 2007).

If \( s^2_u \) is a random variable representing the variance of a random sample of size \( r \) taken from a normal population having the variance \( \sigma^2_u \), then the statistic
\[ \chi_u^2 = \frac{(r-1)s_u^2}{\sigma_u^2} \] (B.3)

has a Chi-square distribution with \( \nu = r - 1 \) degrees of freedom. The detailed functional form of the Chi-square distribution described in equation (B.3) can be expressed as

\[
f\left(\chi_u^2, r-1\right) = \frac{1}{2^{r/2}\Gamma\left(r/2\right)} \left(\frac{(r-1)s_u^2}{\sigma_u^2}\right)^{(r-1)/2} \cdot \exp\left[-\frac{(r-1)s_u^2}{2\sigma_u^2}\right]
\] (B.4)

The expectation of \( \chi_u^2 \) then can be obtained by the following equation, viz., (B.5).

\[
E\left(\chi_u^2\right) = \int \chi_u^2 f\left(\chi_u^2, r-1\right) d\chi_u^2 = r - 1.
\] (B.5)

Since the sample variance \( s_{u, obs}^2 \) is used instead of the population variance \( \sigma_u^2 \), equation (B.4) can be rewritten as:

\[
f\left(\chi_u^2, r-1\right) = \frac{1}{2^{r/2}\Gamma\left(r/2\right)} \left(\frac{(r-1)s_{u, obs}^2}{s_u^2}\right)^{(r-1)/2} \cdot \exp\left[-\frac{(r-1)s_{u, obs}^2}{2s_u^2}\right].
\] (B.6)

Letting \( \chi^2 \) be a vector in which \( \chi_u^2 \) is the \( u \)th element, the vector form of \( \chi^2 \) can be defined as

\[
\chi^2 = \begin{bmatrix} \chi_1^2 & \chi_2^2 & \cdots & \chi_n^2 \end{bmatrix}^T
= \begin{bmatrix} \frac{(r-1)s_1^2}{s_{1, obs}^2} & \frac{(r-1)s_2^2}{s_{2, obs}^2} & \cdots & \frac{(r-1)s_n^2}{s_{n, obs}^2} \end{bmatrix}.
\] (B.7)

Equation (B.2) can have a generalized form as

\[
s^2 = \begin{bmatrix} s_1^2 & s_2^2 & \cdots & s_n^2 \end{bmatrix}^T = X_m v = \begin{bmatrix} X_{1,1} m_v & X_{1,2} m_v & \cdots & X_{1, n} m_v \end{bmatrix}^T
\] (B.8)

where \( X_{uv} \) denotes the \( u \)th row of the design matrix \( X \). By integrating equation (B.8) into equation (B.7), equation (B.7) can be rewritten as

\[
\chi^2 = \begin{bmatrix} \frac{(r-1)X_{1,1} m_v}{s_{1, obs}^2} & \frac{(r-1)X_{1,2} m_v}{s_{2, obs}^2} & \cdots & \frac{(r-1)X_{1, n} m_v}{s_{n, obs}^2} \end{bmatrix}^T = Km_v.
\] (B.9)

where \( K = \begin{bmatrix} \frac{(r-1)X_{1,1}}{s_{1, obs}^2} & \frac{(r-1)X_{1,2}}{s_{2, obs}^2} & \cdots & \frac{(r-1)X_{1, n}}{s_{n, obs}^2} \end{bmatrix}^T \) is a modified form of the design matrix \( X \). The joint probability density function of the vector \( \chi^2 \) then becomes
Development of a new robust design methodology

$$f(\chi^2, F) = \frac{1}{\left(\frac{\Gamma(r/2)}{2^{r/2}}\right)^{n}} \prod_{u=1}^{n} \left(\frac{(r-1)s_u^2}{s_{u \text{ obs}}^2}\right)^{-\frac{r-1}{2}} \cdot \exp\left[-\sum_{u=1}^{n} \frac{(r-1)s_u^2}{2s_{u \text{ obs}}^2}\right]$$  \hspace{1cm} (B.10)

where $F = [r-1 \ r-1 \ \cdots \ r-1]^T$ with $(n - 1)$ degrees of freedom. The prior information of the relationship between the vector of variance and the model parameters $m_v$ is defined as

$$f^{pri}(X_{m_v}) = f(\chi^2, F).$$  \hspace{1cm} (B.11)

Assuming that there is no prior information on the model parameters, $f^{pri}(m)$ in equations (14) and (22) will be a constant, and the posterior information of the model parameters can then be derived as

$$f^{pos}(m_v) = f^{pri}(X_{m_v}) = \frac{1}{\left(\frac{\Gamma(r/2)}{2^{r/2}}\right)^{n}} \prod_{u=1}^{n} \left(\frac{(r-1)s_u^2}{s_{u \text{ obs}}^2}\right)^{-\frac{r-1}{2}} \cdot \exp\left[-\sum_{u=1}^{n} \frac{(r-1)s_u^2}{2s_{u \text{ obs}}^2}\right]$$  \hspace{1cm} (B.12)

Based on Lemma 2, the expectation and covariance matrix of the model parameters $m_v$ for the variance function can be estimated by using the posterior probability density function as detailed in Proposition B.2.

Proposition B.2: Using the assumptions discussed in Proposition B.1, the posterior information of the model parameters can be obtained as

$$\hat{m}_v = \left(K^T K\right)^{-1} K^T F$$  \hspace{1cm} (B.13)

where $K$ is the modified form of the design matrix $X$ as addressed in equation (B.9).

Proof: According to equations (14) and (22), the posterior information of the model parameters can be estimated by substituting the probability density function $f^{pos}(m_v)$ as follows:

$$\hat{m}_v = E(m_v) = \int m_v f^{pos}(m_v) d\!m_v.$$  \hspace{1cm} (B.14)

Through the result of equation (B.11), equation (B.14) will become

$$\hat{m}_v = E(m_v) = \int m_v f(\chi^2, F) d\!m_v.$$  \hspace{1cm} (B.15)

Similar to equation (B.5), the expectation of $\chi^2$ will become

$$E(\chi^2_v) = \int \chi^2_v f(\chi^2_v, F) d\chi^2_v = F$$  \hspace{1cm} (B.16)
or
\[
\int \mathbf{K} \mathbf{m}_v f \left( \mathbf{x}_u^2, \mathbf{F} \right) d\mathbf{x}_u^2 = \mathbf{F} \quad \text{(B.17)}
\]

and
\[
\int \mathbf{m}_v f \left( \mathbf{x}_u^2, \mathbf{F} \right) d\mathbf{x}_u^2 = \left( \mathbf{K}^T \mathbf{K} \right)^{-1} \mathbf{K}^T \mathbf{F}. \quad \text{(B.18)}
\]

Using equations (B.15) and (B.18), the expectation of the model parameters \( \mathbf{m} \), can be expressed as
\[
\mathbf{\hat{m}}_v = \left( \mathbf{K}^T \mathbf{K} \right)^{-1} \mathbf{K}^T \mathbf{F}. \quad \text{(B.19)}
\]

Once the vector of model parameters can be estimated by the results of Proposition B.1, the variance model can finally be formulated as follows:
\[
\hat{\sigma}^2_{1p}(x) = \mathbf{x}^T \mathbf{\hat{m}}_v \quad \text{(B.20)}
\]

where the vector of model parameters \( \mathbf{\hat{m}}_v \) can be estimated by equation (B.13).