Desirability function approach: A review and performance evaluation in adverse conditions

Nuno R. Costa a, c, *, João Lourenço a, b, Zulema L. Pereira c

Abstract

Adverse conditions in terms of quality of predictions and robustness are simulated to evaluate the ability of desirability-based methods for yielding compromise solutions with desired response’s properties. The method’s solutions are assessed at optimal variable settings with respect to bias, quality of predictions and robustness through optimization measures, and the usefulness of those measures to select the compromise solution is evaluated. Three examples with different features in terms of responses variance are used and the performance of various analysis methods is compared. Results show that a less sophisticated desirability-based method can compete with other methods designed to perform well under adverse conditions and that the optimization measures justify its use in real life problems.

1. Introduction

A widely used methodology for developing, improving and optimizing systems (process and product), the so-called response surface methodology (RSM), consists of the following general phases: 1) screening: experiments are designed with the purpose of discovering the vital few control factors that cause statistically significant effects of practical importance for the goal of the study; 2) modeling: experiments are designed with the purpose of modeling the quality characteristic of interest (response) as a function of control factors; 3) optimization: response model is analyzed to determine the variable setting at which optimum conditions of system property are achieved.

This methodology has inherent sequential experimentation strategy that, if the technical and non-technical issues in the experimental phases are properly managed, leads to a high level of system knowledge. For an understanding of the assumptions and conditions necessary to successfully apply RSM the reader is referred to Refs. [1–3].

This article focuses on the optimization phase and, in particular, on the optimization of multiple responses. These problems are usual in various fields of science and often involve incommensurate and conflicting responses that must be in some sense optimized simultaneously, because their separate analysis may result in incompatible solutions. This has been a much researched subject and a strategy widely used in the RSM framework consists of converting the multiple responses into a single one by combining the individual responses into a composite function followed by its optimization. Although the desirability function and loss function approaches are popular among practitioners, other approaches have been used for optimizing multiple responses. Among them are those based on Compromise Programming [4], Goal Programming [5, 6], Inspection of contour plots [7], Physical Programming [8, 9], Probability-based [10], Performance Index [11, 12], Neural Networks [13, 14], and Vectorial Optimization [15]. All these approaches have their own merits. However, the lack of recommendations for proper use, unavailability of the algorithms employed, and (mathematical/statistical) complexity of some approaches and methods are major reasons by which they are of little practical use or appealing to practitioners, in particular to non-statisticians.

The less sophisticated desirability-based methods are easy to understand, easy to use, and flexible for incorporating the decision-maker’s preferences (weights or priorities assigned to responses). Moreover, the most popular desirability-based method, the Derringer and Suich’s method [16] or modifications of it [17] are available in many data analysis software packages. This method has been extensively used in practice, namely in chemometrics, such as illustrated in Refs. [18–22]. However, the analyst needs to specify values to four types of parameters (shape factors and weights) to use the so-called Derringer and Suich’s method. This is not a simple task...
and impacts on the optimal variable settings. So, other less known and used methods that require less subjective information from the user and can yield effective compromise solutions are particularly welcome, namely to non-statistician practitioners who are using statistics more than ever before [23].

This article provides a review on existing desirability-based methods and aims at evaluating the ability of two less sophisticated ones along with appropriate optimization performance measures to give desirable response’s properties at optimal variable settings, namely low bias (responses deviation from target) and low variance. This is illustrated with three case studies where adverse conditions in terms of responses variance are simulated. In the first simulated scenario the response’s models differ in terms of the quality of predictions (variance due to uncertainty in the regression coefficients). In the second simulated scenario the response’s models are characterized by unequal sensitivity to uncontrollable variables (robustness) and high quality of predictions. In the third one, both the quality of predictions and robustness are low.

The remainder of the article is organized as follows: Section 2 provides a review on desirability-based methods; optimization criteria and measures to evaluate the compromise solutions in terms of response’s bias, quality of predictions and robustness are introduced in Section 3; Section 4 characterizes the simulated scenarios and includes three examples to illustrate the feasibility of the proposed approach; Section 5 includes the discussion of results and Section 6 presents the final conclusions.

2. Literature review

Desirability function-based approach consists of converting the estimated response models \( \hat{y} \), which usually are second order models, into individual desirability functions \( d \) that are then aggregated into a composite function \( D \). This function is usually a geometric or an arithmetic mean, which will be maximized or minimized, respectively.

Next subsections provide a chronological review on the existing desirability-based methods, which are separated into two groups of different mathematical/statistical sophistication levels. The first group includes the less sophisticated approaches, while the second group includes the more sophisticated approaches.

2.1. Less sophisticated approaches

This subsection reviews the methods proposed in Refs. [16,17] and [24–28]. These methods do not require from the analyst high mathematical and statistical expertise as they are easy to understand and can be easily implemented in Excel®.

2.1.1. Derringer and Suich [16]

These authors generalized the exponential type transformations proposed in Ref. [29] to transform the responses into desirability functions and developed more flexible desirability functions. Derringer and Suich proposed individual desirability functions based on three response types as follows:

A. Nominal-The-Best (NTB): the value of the estimated response is expected to achieve a particular target value \( T \). For this response type, the individual desirability function is defined as

\[
d = \begin{cases} 
\left( \frac{\hat{y} - L}{T - L} \right)^r, & L \leq \hat{y} \leq T \\
\left( \frac{\hat{y} - U}{T - U} \right)^r, & L \leq \hat{y} \leq T \\
0, & \text{otherwise}
\end{cases}
\]  

(1)

where \( s \) and \( t \) are user-specified parameters \((s, t>0)\) that allow practitioners to specify the shape of \( d \). The individual desirability function transforms the response variable into a range of values between 0 and 1, where 1 is most favorable. In the NTB case, \( d = 1 \) for \( \hat{y} = T \), and \( d = 0 \) for \( \hat{y} < L \) or \( \hat{y} > U \), where \( U \) is the upper and \( L \) is the lower specification limit of the response.

B. Larger-The-Best (LTB): The value of the estimated response is expected to be larger than a lower bound \( \hat{y} \). For this response type, the individual desirability function is defined as

\[
d = \left( \frac{\hat{y} - L}{U - L} \right)^r, \quad L \leq \hat{y} \leq U
\]  

(2)

where \( r \) is a user-specified parameter \((r>0)\). In the LTB case one assumes that it is possible to establish a finite target \( U \) such that \( d = 1 \) for \( \hat{y} \geq U \) and \( d = 0 \) for \( \hat{y} \leq L \).

C. Smaller-The-Best (STB): The value of the estimated response is expected to be smaller than an upper bound \( \hat{y} \). For this response type, the individual desirability function is defined as

\[
d = \left( \frac{\hat{y} - U}{T - U} \right)^r, \quad L \leq \hat{y} \leq U
\]  

(3)

under the assumption that it is possible to establish a finite target \( L \) such that \( d = 1 \) for \( \hat{y} > L \), and \( d = 0 \) for \( \hat{y} \geq U \).

The values assigned to \( s, t \) and \( r \) allow changing the shape of \( d \). A large value for \( s, t \) or \( r \) implies that the individual desirability value is very low unless the response gets very close to its target value. This means that the higher their values are, the greater the importance of that response type.

This article provides a review on existing desirability-based methods; optimization criteria and measures to evaluate the compromise solutions in terms of response’s bias, quality of predictions and robustness are introduced in Section 3; Section 4 characterizes the simulated scenarios and includes three examples to illustrate the feasibility of the proposed approach; Section 5 includes the discussion of results and Section 6 presents the final conclusions.
where \( f(\hat{y}) = A + B\hat{y} + C\hat{y}^2 + D\hat{y}^3 + E\hat{y}^4 \), \( \delta = (U - L)/50 \) and \( A, B, C, D, E, a_0, b_0, a_1 \) and \( b_1 \) are constants. The composite function is given by Eq. (4).

The major drawback of this method is that it cannot be applied in problems with STB- and LTB-type responses, which are usual in practice.

2.1.3. Wu and Hamada [25]

Wu and Hamada suggested individual desirability functions in the form of exponential functions, which for NTB-type response are defined as

\[
d = \begin{cases} 
\exp(-k_1(\hat{y} - T_i^F)), & -\infty < \hat{y} \leq T_i \\
\exp(-k_2(\hat{y} - T_i^L)), & T_i < \hat{y} < \infty 
\end{cases}
\]

where \( k_1 \) and \( k_2 \) are scale constants and \( s \) and \( t \) are shape parameters.

For STB- and LTB-type responses the individual desirability functions are defined by Eqs. (7–8), respectively.

\[
d = \exp(-k_1(\hat{y} - \phi_i^F)), \quad \phi_i \leq \hat{y} < \infty
\]

\[
d = \begin{cases} 
1 - \exp(-k_1 \hat{y}) & \quad L \leq \hat{y} < \infty \\
0 & \quad \hat{y} < L
\end{cases}
\]

where \( k_1 \) is a scale constant, \( \phi \) represents the target value for STB-type responses, and \( r \) is a shape parameter.

As an optimization criterion, the authors proposed to maximize the global desirability function \( D \)

\[
D = (d_1)^{w_1} (d_2)^{w_2} \cdots (d_p)^{w_p}
\]

with \( 0 < w_i < 1 \) and \( w_1 + w_2 + \ldots + w_p = 1 \). Note that this method requires a number of user-specified parameters larger than the so-called Derringer and Suich’s method, which is not an appealing feature.

2.1.4. Kim and Lin [26]

These authors proposed to maximize the overall minimal level of satisfaction with respect to all the responses. To this purpose an exponential desirability functional form that can generate a rich variety of shapes by adjusting its parameters is suggested. It does not require any assumptions regarding the form or degree of the estimated response models, is robust to the potential dependences between response variables, and considers the relative priority of responses. The individual desirabilities are defined as

\[
d = \begin{cases} 
\exp(k - \exp(k|z|)), & t \neq 0 \\
1 - |z|, & t = 0
\end{cases}
\]

where \( k \) is a constant \((-\infty < k < \infty)\) and \( z \) is a standardized parameter representing the distance of the estimated response from its target in units of the maximum allowable deviation. This parameter depends on the response type and is defined as

\[
z = \begin{cases} 
\frac{\hat{y} - T_i}{T_i - L_i} & \text{for NTB} \\
\frac{\hat{y} - \hat{y}_{min}}{\hat{y}_{max} - \hat{y}_{min}} & \text{for STB with } \hat{y}_{min} \leq \hat{y} \leq \hat{y}_{max} \\
\frac{\hat{y} - \hat{y}_{max}}{\hat{y}_{max} - L_i} & \text{for LTB}
\end{cases}
\]

where \( \hat{y}_{max} \) and \( \hat{y}_{min} \) represent the maximum and minimum values of the estimated response, respectively. Eq. (11) ranges between \(-1 \) and \( 1 \) for NTB-type responses and between \( 0 \) and \( 1 \) otherwise. This modeling approach can take into account the level of responses predictive ability through the adjustment of the parameter \( k \). In this case, it will be defined as

\[
k' = k + \left(1 - R^2\right)^{k_{\max} - k}
\]

where \( k_{\max} \) is a sufficient large value of \( k \) such that \( d(z) \) with \( k_{\max} \) is extremely concave and thus has virtually no effect in the optimization process. \( R^2 \) is the coefficient of determination.

The composite function to maximize the overall minimal level of satisfaction with respect to all the responses is

\[
D = \max \min (d_1, d_2, \ldots, d_p)
\]

This maximin scheme has a serious disadvantage: it only considers the response with the lowest degree of satisfaction so the degrees of satisfaction associated with all the other responses are ignored, which may lead to an unreasonable decision in some cases. As an example, the approach would suggest a solution with satisfaction levels equal to \((0.6, 0.6, 0.6, 0.6)\) instead of \((0.9, 0.9, 0.9, 0.58)\).

2.1.5. Ch’ng et al. [27]

Minimization of an arithmetic mean is the proposal of Ch’ng et al. for finding the optimal variable settings under the assumption of normality and homogeneity of error variances. They proposed individual desirability functions of the form

\[
d = \frac{2\hat{y} - (U + L)}{U - L} + 1 = \frac{2\hat{y} - 2L}{U - L} + \frac{-2L}{U - L} = \frac{m\hat{y} + c}{2}
\]

with \( 0 \leq d \leq 2 \). As an optimization criterion, the authors minimize a composite function defined as

\[
D = \left(\sum_{i=1}^{p} w_i |d_i - d(T_i)|\right)/p
\]

where \( d(T_i) \) represents the value of the \( i \)-th individual desirability function for \( \hat{y} \) equal to the response target \( (\hat{y} = T_i) \), \( w_i \) represents the weight (importance or priority) assigned to response \( i \), \( p \) is the number of responses, and \( \sum_{i=1}^{p} w_i = 1 \).

Distinctive features of this method are the following: individual desirabilities do not have any breakpoint; can be easily understood and implemented; and subjective information required from the analyst is minimal. Moreover, the number of user-specified parameters (weights) is just equal to the number of responses, which is an appealing feature. The theoretical drawback is that it does not consider the variances and correlations existing among the responses.

2.1.6. Wu [28]

In contrast to previous authors, Wu [28] proposed individual and composite desirability functions that consider the variances and correlations among the responses. The individual desirability function of the \( i \)-th response is defined as

\[
d_{\hat{y}_i} = \exp(-F_{\hat{y}_i})
\]

where \( F_{\hat{y}_i} = \left[k_i \left(\hat{y}_i - T_i\right)^2 + \hat{y}_i^2\right]^{1/2} \), \( k_i \) is the loss coefficient and \( \hat{y}_i^2 \) is the estimated variance model for the \( i \)-th response \( (\hat{y}_i) \). The correlated desirability function \( (d_{\hat{y}_i}, i \neq j) \) is defined as

\[
d_{\hat{y}_i} = \begin{cases} 
1 - \exp(-F_{\hat{y}_i}) & \quad F_{\hat{y}_i} < 0 \\
\exp(-F_{\hat{y}_i}) & \quad F_{\hat{y}_i} \geq 0
\end{cases}
\]
where $F_g = k_{0g} \left( \alpha_k \beta_k + (\bar{y}_g - T_i) \left( \bar{y}_g - T_j \right) \right)$, $k_{0g}$ is the correlated loss coefficient and $\alpha_k$ is the correlation coefficient of responses $y_i$ and $y_j$. As optimization criterion they maximize the global desirability shown in Eq. (18).

$$D^* = \left[ \prod_{i=1}^{r} \prod_{j=1}^{c} d_i d_j \right]^{1/s}$$ (18)

The major drawback of this method is the number and computation of the user-specified parameters (denoted by $k_i$ and $k_j$), which represent the loss coefficient of the $i$-th response and correlated loss coefficient of responses $i$ and $j$ to which the analyst needs to assign values. These parameters play a decisive role in the optimization process. However, its number is always larger than the number of responses and the information necessary to compute them is not easy to define or readily available. For details on the computation of the loss and correlated loss coefficients the reader is referred to Wu and Chyu [31].

2.2. More sophisticated approaches

This subsection provides a brief summary of some desirability-based methods that require from the user a higher level of expertise on mathematics and statistics, which may limit their use in practice. It includes the methods proposed in Refs. [32–37].

2.2.1. Ribardo and Allen [32]

Arguing that there is often an interaction between the mean and standard deviation with regard to their effects on process yield and therefore on profitability, Ribardo and Allen [32] proposed a desirability-based method that explicitly accounts for the combined effect of the mean and dispersion of responses to determine the steepness of the desirability functions and to regulate how far inside the limits the mean should be positioned. These authors employ a criterion that is based on estimates of the process yield, namely on the predicted fraction of conforming parts. However, to preserve the interpretation of the desirability functions as the yield associated to responses, the individual and composite desirabilities become hard to interpret and use.

2.2.2. Ortiz et al. [33]

These authors show that if multiresponse optimization problems grow even moderately in either the number of factors or the number of responses, conventional optimization algorithms may fail to find the optimal variable settings. Thus they proposed an approach that uses a genetic algorithm in conjunction with an unconstrained desirability function, enabling the algorithm to differentiate between far-from-feasible and nearly feasible solutions. Delineation is accomplished by incorporating a penalty function proportional to the magnitude of the constraint violation into the overall desirability value of each design space point.

2.2.3. Pasandideh and Niaki [34]

To integrate the desirability function and simulation approach with a genetic algorithm was the proposal of Pasandideh and Niaki, which consists in: modeling the multiresponse problem by desirability function models; generating the required input data from a simulated system; optimizing the composite function through a genetic algorithm. Four genetic algorithms that are different in structure, especially in controlling the stochastic nature of the problem, are compared and they concluded that no statistical difference exists between them.

2.2.4. Lee and Kim [35]

A new type of desirability function is proposed by these authors. The so-called expected desirability function is defined as the average of the conventional desirability values based on the probability distribution of the predicted response variable. It considers the dispersion effects and location effects of the responses, does not require a separate desirability function for the standard deviation response, and can be used for (moderately) asymmetric NTB-type responses.

2.2.5. Das and Sengupta [36]

In situations where customer perceptions are used as metrics/measurements of critical performance characteristics of the process or product, the scales usually take values between $-k$ and $+k$ for some characteristic(s) and so the individual and composite desirability functions above reviewed are not appropriate. As alternative, Das and Sengupta proposed desirability functions that either do not reflect or avoid using the negative and zero values, because the negative exponential transformation used to modify Gatza’s desirability function can accommodate such values.

2.2.6. He et al. [37]

According to these authors, most of the recent work in industrial optimization does not deal with the quality of predictions and most practitioners follow up optimization calculations without considering this responses property. So He et al. [37] proposed a creative procedure to define individual desirability functions that account for quality of predictions (QoP) and a composite desirability function that makes balance between QoP and bias.

3. Proposed approach

Progress in research on desirability functions has been focused on flexibility, minimization of the impact in decision-making of certain types of dependencies caused by data limitations, consideration of response’s variance, and use of genetic algorithms. Approaches that require minimum subjective information from the user and are easy to understand and implement are especially appealing to practitioners and useful if they provide compromise solutions with desired response’s properties. Therefore, we evaluated the ability of two less sophisticated methods along with optimization performance measures to yield compromise solutions with respect to desired response’s properties under adverse variance conditions. The goal is to identify the best performer, if at all possible. The methods under evaluation are those of Ch’ng et al. [27] and Wu [28].

Ch’ng et al.’s method is adapted to accommodate LTB- and STB-type responses by using $d(U_i)$ and $d(L_i)$ in Eq. (15) instead of $d(T_i)$, under the assumption that it is possible to establish specification limits $U$ and $L$ to those responses. If these limits are not readily available, it is reasonable to use the maximum and minimum values of the respective response model.

Ch’ng et al.’s method is not designed to take into account the responses variance levels and correlation information, so it may lead to less reasonable solutions when the response’s correlation and variance due to low QoP and uncontrollable factors are important issues in practice. To minimize this problem, the estimated variance models in addition to estimated mean models for the responses are considered in Eq. (15) and then optimization measures are used to help the analyst evaluate and select a compromise solution.

Seemingly Unrelated Regression (SUR) method is used to estimate the regression models. This method can lead to a more precise estimate of the control factors setting, in particular if responses are correlated [38,39]. This is corroborated in Ref. [10] where it is showed that SUR modeling can make a noticeable difference.

3.1. Optimization performance measures

The solutions obtained from the methods proposed in Refs. [27,28] are assessed in terms of bias, quality of predictions, and robustness through the optimization measures proposed in Ref. [40].
The bias is assessed with an optimization measure that considers
the response types, response’s specification limits and response
deviations from target. This measure, named cumulative bias ($B_{cum}$),
is defined as

$$B_{cum} = \sum_{i=1}^{p} W_i |\hat{y}_i - 0_i|$$

(19)

where $\hat{y}_i$ represents the value of the $i$-th estimated response at
“optimal” variables setting, $\theta_i$ is the response target value and $W_i$ is
a parameter that takes into account the specification limits and
response type of the $i$-th response. This parameter makes responses
dimensionless and is defined as follows: $W = 1/(U - L)$ for STB and
LTB-type responses; $W = 2/(U - L)$ for NTB-type response.

The cumulative bias gives an overall result of the optimization
process instead of focusing on the value of a single response, which
prevents the analyst from taking unreasonable decisions in some
cases, such as what Kim and Lin [26] showed. Nevertheless, the bias of
each response is also available from Eq. (19).

To assess method’s solutions in terms of quality of predictions is
used a measure defined as follows:

$$QoP = \text{trace} \left[ \phi \left( x^T \Sigma Q^{-1} x \right)^{-1} \phi^T \right] = \text{trace} \left[ \phi \sum \hat{d}^2(x) \right]$$

(20)

where $x_i$ is the subset of independent variables consisting of the $Kx1$
vector of regressors for the $i$-th response with $N$ observations on $K_i$
regressors for response, $x$ is an $NpxK$ block diagonal matrix and
$Q = \sum \Theta_{hi}$. An estimate of $\Sigma$ is $\hat{\Sigma} = \hat{\Theta} \hat{\Theta}^T / N$, where $\hat{\Theta}$ is the residual
vector from the OLS (Ordinary Least Squares) estimation of the $i$-th response;
$\Theta$ is an identity matrix and $\Theta$ represents the Kronecker
product. The matrix $\sum \hat{d}^2(x)$ is calculated for $x_i$ equal to optimal settings and
to make it dimensionless this matrix is multiplied by matrix $\phi$, whose
diagonal and non-diagonal elements are $\phi_i = 1/(U_i - L_i)^2$ and
$\phi_{ij} = 1/(U_i - L_i)(U_j - L_j)$ for $i \neq j$, respectively.

This measure is defined under the assumption that Seemingly
Unrelated Regression (SUR) method is used to estimate the regression
models. If the Ordinary Least Squares is used, the reader is referred to
Ref. [41] where variants of Eq. (20) are presented for the cases of
regression models with equal and different terms.

The robustness is assessed by

$$\text{Rob} = \text{trace} \left[ \phi \hat{\Theta} \hat{\Theta}^T \right]$$

(21)

where $\hat{\Theta}_{hi}$ represents the variance–covariance matrix of the “true”
responses at “optimal” variable setting. Note that replications of the
experimental runs are required to calculate the variance–covariance
matrix and matrix $\phi$ only considers the specification limits assigned to
responses variance. Although the replications increase the time and cost
of experimentation, they may provide significant improvements in
robustness that overbalance or at least compensate the time and cost
spent.

The optimization measures defined in Eqs. (19–21) do not exclude
other measures from also being used. As concerns their results, the
lower the values are, the better the compromise solution will be. In
practice, the presented measures take values greater than or equal to
zero, being the most favorable the zero value.

3.2. Select a compromise solution

Choosing the best compromise solution from a set of competing
alternatives in multiresponse problems is not a trivial task. In the RSM
framework the general procedure consists of selecting that solution
based on the output (maximum or minimum value) of the objective
function used. However, the widely used global desirability and
expected loss functions give inconsistent and incomplete information
to the analyst in terms of the merit of the final solution, namely with
respect to desired response’s properties. If the analyst only focuses on
the result of the composite function used for making decisions he/she
may ignore a solution of interest, because that value depends on the
weights assigned to responses and a satisfactory solution may have
less favorable desirability values. Such as what Costa et al. [40]
showed, the composite functions yield different results in cases where
the solutions are equal or have slight changes in the response values
due to the different weights or priorities assigned to responses. This is
a relevant shortcoming that the analyst must be aware of and may
overtake by using the aforementioned optimization measures,
because they do not depend on the priorities assigned to responses.
However, another problem arises in choosing a compromise solution
because some of them may be better with respect to bias but worse
with respect to variance (QoP and/or Rob). In these instances, unless
technical or economical reasons arise to justify another option, the
cumulative value of $B_{cum}$, QoP and Rob measures, denoted by Cum, is
considered in this article to select the compromise solution.

4. Examples

Myers and Montgomery [42] reported an experimental study
where the objective was to maximize the conversion of a polymer and
minimize the thermal activity by setting the variables reaction time ($x_1$),
reaction temperature ($x_2$), and amount of catalyst ($x_3$) at
appropriate levels. The ranges for the conversion ($y_1$) and thermal
activity ($y_2$) responses are [80, 100] and [55, 60], respectively.
Assuming that $y_1$ is a LTB-type response, its target value is set equal
to 100; $y_2$ is a NTB-type response and its target is set equal to 75.5.

This example supports the design of three case studies to evaluate
and compare the performance of several analysis methods, which were
coded in Matlab®. Adverse variance conditions are simulated for each
case study as follows: Case 1 — low quality of predictions; Case 2 — high
quality of predictions and unequal robustness; and Case 3 — low quality
of predictions and unequal robustness.

Cases 2–3 assume unequal robustness, which implies that, in
addition to estimated models for the mean response ($\hat{\mu}_1$ and $\hat{\mu}_2$,
respectively), replicates are needed to estimate the variance models
for polymer conversion ($\hat{\sigma}_1$) and thermal activity ($\hat{\sigma}_2$). Based on
the central composite design with six axial and six center points ran by
Myers and Montgomery (see Table 1), five replicates ($i=1,2,...,5$)
are generated, like in Ref. [43], using the model

$$\begin{pmatrix} y_1(x) \\ y_2(x) \end{pmatrix} = \begin{pmatrix} y_1(x) \\ y_2(x) \end{pmatrix} + e_i$$

(22)

where $y_1(x)$ and $y_2(x)$ represent the response values presented in Ref.
[42], and the error term $e_i$ is defined as follows:

$$e_i \sim N \left[ 0, \begin{pmatrix} \sigma_{11}(x) & \sigma_{12}(x) \\ \sigma_{21}(x) & \sigma_{22}(x) \end{pmatrix} \right]$$

(23)

with $i=1,2,...,5$, $\sigma_{11}(x) = \exp(3 - x_1^2 - 3x_1^2)$, $\sigma_{22}(x) = \exp(2 - 2x_1^2 - x_3^2)$, and $\sigma_{12}(x) = \sigma_{21}(x) = 0.03x_1(x)\sigma_{22}(x)$. These models make the
responses variance due to uncontrollable variables larger for
operation conditions near the origin, that is, the closer the operating
conditions for $x_1$ and $x_3$ are to the origin, the worst the responses
robustness will be. The responses models are fitted using SUR method
and the range for the input variables is $-1.682 \leq x_i \leq 1.682$, $j=1,2,3$.

Ching et al.’s and Wu’s methods are evaluated in three case studies
and the feasibility of their solutions compared with those of methods
designed to consider the quality of predictions and robustness. The
method proposed by Vining [41], which allows specifying the
directions of economic importance (weights or priorities to responses)
while considering the variance–covariance structure of the estimated
responses (quality of predictions) is used in Case 1. The method
proposed by Pignatiello [44], which puts emphasis on improving
robustness, by considering the variance–covariance structure of the “true” responses and minimizing bias, is used in Case 2. In Case 3, the method proposed by Ko et al. [43] is used because it is designed to consider both quality of predictions and robustness in addition to bias. This method is built on the methods presented in Refs. [41] and [44], and yields better results than these two methods when both quality of predictions and robustness are important issues in practice [43].

The expected loss function proposed in Ko et al. [43] includes the expected losses of Vining [41] and Pignatiello [44] as special cases and is defined as follows:

$$E[L(y(x), \theta)] = (y(\hat{x}) - \theta)^\prime C(y(\hat{x}) - \theta) + \text{trace}[C(\sum y(x))] + \text{trace}[C(\sum x)]$$

(24)

where the first term is an imposed penalty due to bias, the second term is the imposed penalty due to poor quality of predictions, and the third term is the imposed penalty due to poor robustness. The first and second terms constitute the Vining’s expected loss function; the first and third terms constitute the Pignatiello’s expected loss function. The variance–covariance matrix of the predicted responses at x is denoted by $\sum y(x)$ while $\sum x(x)$ is the variance–covariance matrix of the responses at x. C is a cost matrix and their elements represent the costs of non-optimal variable setting. If C is a diagonal matrix then each element represents the relative importance assigned to the respective response, that is, the penalty (cost) incurred for each unit of response value deviates from its optimum. If C is a non-diagonal matrix, the off-diagonal elements represent additional costs incurred when pairs of responses are off-target simultaneously.

4.1. Case 1 — low quality of predictions

Some insignificant regressors ($x_1$, $x_2$, $x_3$, $x_4$, $x_5$, $x_6$) were added to the model fitted to thermal activity so that the predicted response has a lower quality of predictions when $x_i$ values are farther from the origin. The estimated models are:

$$\hat{y}_1 = 81.0943 + 1.0290x_1 + 4.0426x_2 + 6.2060x_3 - 1.8377x_1^2 + 2.9455x_2^2 - 5.2036x_3^2 + 2.1250x_1x_2 + 11.3750x_1x_3 - 3.8750x_1x_3$$

$$\hat{y}_2 = 59.8505 + 3.5855x_1 + 0.2547x_2 + 2.2312x_3 + 0.8360x_1^2 + 0.0742x_2^2 + 0.0565x_3^2 - 0.3875x_1x_2 - 0.0375x_1x_3 + 0.3125x_2x_3$$

According to Vining [41] the objective of this problem is to achieve a value for $\bar{y}$ as large as possible, but the product must be sellable. This occurs when the value of thermal activity is inside the specification limits, although the ideal situation is to have $\bar{y}$ on target. To bring thermal activity closer to the target, Vining assigned a larger weight to $\bar{y}$ than to $\bar{y}$. At this condition his method yielded a

<table>
<thead>
<tr>
<th>Solution</th>
<th>Weights</th>
<th>$x_i$</th>
<th>$\bar{y}_1$, $\bar{y}_2$</th>
<th>Result</th>
<th>Bes</th>
<th>QoP</th>
<th>Cum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vining</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V11</td>
<td>(0.0, 0.00, 50.0)</td>
<td>(-0.4898, 1.68, -0.578)</td>
<td>95.19, 57.67</td>
<td>$E_{\text{cum}} = 98.41$</td>
<td>0.548</td>
<td>0.066</td>
<td>0.614</td>
</tr>
<tr>
<td>V21</td>
<td>(0.01, 0.025, 0.50)</td>
<td>(-0.3781, 1.68, -0.4871)</td>
<td>95.24, 58.16</td>
<td>$E_{\text{cum}} = 3.73$</td>
<td>0.742</td>
<td>0.059</td>
<td>0.801</td>
</tr>
<tr>
<td>V31</td>
<td>(0.01, 0.04, 0.30)</td>
<td>(-0.3716, 1.67, -0.4243)</td>
<td>94.18, 58.39</td>
<td>$E_{\text{cum}} = 0.50$</td>
<td>0.938</td>
<td>0.049</td>
<td>0.987</td>
</tr>
<tr>
<td>V41</td>
<td>(0.02, 0.01, 0.80)</td>
<td>(-0.4776, 1.68, -0.5654)</td>
<td>95.19, 57.73</td>
<td>$E_{\text{cum}} = 1.68$</td>
<td>0.572</td>
<td>0.005</td>
<td>0.617</td>
</tr>
<tr>
<td>V51</td>
<td>(0.01, 0.04, 0.80)</td>
<td>(-0.4299, 1.526, -0.4871)</td>
<td>92.95, 57.97</td>
<td>$E_{\text{cum}} = 1.17$</td>
<td>0.892</td>
<td>0.046</td>
<td>0.938</td>
</tr>
<tr>
<td>Ch'ng et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>(0.22, 0.78)</td>
<td>(-0.5434, 1.68, -0.5954)</td>
<td>91.25, 57.50</td>
<td>D = 0.053</td>
<td>0.479</td>
<td>0.068</td>
<td>0.547</td>
</tr>
<tr>
<td>C21</td>
<td>(0.89, 0.11)</td>
<td>(0.0221, 1.68, -0.2019)</td>
<td>96.13, 60.00</td>
<td>D = 0.227</td>
<td>1.387</td>
<td>0.049</td>
<td>1.436</td>
</tr>
<tr>
<td>C31</td>
<td>(0.86, 0.14)</td>
<td>(-1.68, 1.68, -1.058)</td>
<td>98.03, 55.00</td>
<td>D = 0.155</td>
<td>1.197</td>
<td>0.243</td>
<td>1.440</td>
</tr>
<tr>
<td>C41</td>
<td>(0.01, 0.99)</td>
<td>(-0.6576, 1.68, -0.5163)</td>
<td>94.05, 57.50</td>
<td>D = 0.002</td>
<td>0.496</td>
<td>0.070</td>
<td>0.566</td>
</tr>
<tr>
<td>Wu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W11</td>
<td>(0.01, 0.01, 0.30)</td>
<td>(-0.5436, 1.68, -0.5982)</td>
<td>95.21, 57.50</td>
<td>D$^2$ = 0.891</td>
<td>0.479</td>
<td>0.069</td>
<td>0.548</td>
</tr>
<tr>
<td>W21</td>
<td>(0.08, 0.01, 0.40)</td>
<td>(-0.2125, 0.346, 0.3163)</td>
<td>82.64, 60.00</td>
<td>D$^2$ = 0.000</td>
<td>2.736</td>
<td>0.013</td>
<td>2.749</td>
</tr>
</tbody>
</table>
solution denoted by V11 in Table 2. However, a solution with \( B_{\text{cum}} \)-QoP and Cum values in close agreement with V11, denoted by V41, was achieved without the need of assigning a large weight to \( \hat{\mu}_2 \), which justifies the difference in their expected loss (\( E_{\text{Loss}} \)) values.

The set of solutions yielded by Vining’s method is presented in Fig. 1. None of the solutions have \( \hat{\mu}_2 \) on target and those with better \( \hat{\mu}_2 \) values have worse \( \hat{\mu}_1 \) values and vice-versa. Except V11, all the other solutions have \( \hat{\mu}_2 \geq 0.73 \) with 93.77 < \( \hat{\mu}_1 < 95.99 \).

These results were achieved by testing the loss coefficients as follows: \( c_1 = \{0.01, 0.02, \ldots , 0.2\} \); \( c_2 = \{0.3, 0.4, \ldots , 0.8\} \). This set of values include the weights used by Vining [41], \( c_1 = 0.1 \), \( c_2 = 0.25 \) and \( c_2 = 0.5 \), which lead to the solution V21. This solution gives worse QoP and \( E_{\text{Loss}} \) values than V31 and V51, which are the solutions with the lowest \( E_{\text{Loss}} \) and QoP values, respectively. Moreover, V21 gives worse \( E_{\text{Loss}} \), \( B_{\text{cum}} \) and Cum values than V41. These situations, like that one highlighted at the end of the first paragraph, points out how critical is to set values to loss coefficients.

In the Ch’ng et al.’s method, the change or increment for the weights was set equal to 0.01 units, under the constraint that the sum of \( c_i \) equals one. Most solutions (76 out of 99) have the same \( x_1 \) values. This compromise solution, denoted by C11 in Table 2, gives the lowest Cum value and is characterized for having \( \hat{\mu}_1 \) on target. In the other solutions, a degradation of \( \hat{\mu}_1 \) value occurs, that is, \( \hat{\mu}_1 \) is moved to the specification limit, while \( \hat{\mu}_2 \) is closer to the target. Examples are the solutions C21 and C31, which are displayed in Fig. 1 rounded by a circle. Solution C21 gives the lowest QoP value among the solutions yielded by Ch’ng et al.’s method, and its value is equal to that of V31.

The solution C41 gives the lowest composite function (\( D \)) value, although it is slightly worse than C11 with respect to Cum. Note that C11 gives a better Cum value than all the solutions yielded by Vining’s method. The solutions of Ch’ng et al.’s method are displayed in Fig. 1 as asterisks, but some of them, like the C11, are not visible because they are hidden by Wu’s method results, which are displayed as triangles. In fact, some solutions (146 out of 480) yielded by Wu’s method, where the response weights were set equal to those used in the Vining’s method, are identical or in close agreement with C11. These solutions give the largest desirability value and present a Cum value lower than all the solutions yielded by Vining’s method. The remaining solutions have \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) close or equal to the upper specification limit. Examples are the solution rounded by a circle and the solution denoted by W21, which give an unacceptable \( D \) value. Solutions like W21 (318 out of 480, with 2.52 < Cum < 2.84) are not shown in Fig. 1, because only solutions with \( x_3 = 1.682 \) are displayed.

### 4.2. Case 2 — high quality of predictions and unequal robustness

Besides the two estimated models for the mean response (\( \hat{\mu}_1 \), \( i = 1, 2 \)), two models are fitted for standard deviation (\( \sigma \)) so that responses robustness is considered. The models are as follows:

\[
\hat{\mu}_1 = 81.7863 + 0.8234 x_1 + 4.1361 x_2 + 6.2127 x_3 + 1.9954 x_1^2 + 2.8346 x_2^2 - 5.5585 x_3^2 + 1.6981 x_1 x_2 + 11.4114 x_1 x_3 - 3.8694 x_2 x_3
\]

\[
\hat{\sigma}_1 = 4.4903 - 1.2366 x_1 - 0.8904 x_2^2 - 1.6466 x_3^2
\]

\[
\hat{\mu}_2 = 60.7977 + 3.5894 x_1 + 2.1794 x_2
\]

\[
\hat{\sigma}_2 = 2.0583 - 0.6512 x_1^2 - 0.7015 x_3^2
\]

The specification limits and targets for \( \hat{\sigma}_1 \) and \( \hat{\sigma}_2 \) are: \( \hat{\sigma}_1 \leq 10 \) and \( \hat{\sigma}_2 \leq 5 \) with targets equal to zero. The response weights and increments are the same as in Case 1.

Pignatiello solutions with the lowest \( E_{\text{Loss}} \), Cum and Rob values are denoted in Table 3 by P12, P22 and P32, respectively. As regards the Cum value, the best Pignatiello solution (P22), delimited by a hexagon in Figs. 2 and 3, is in close agreement with the best ones yielded by Ch’ng et al.’s method (C12 and C22). These solutions also have the lowest \( B_{\text{cum}} \) value. The solution P32 has a \( B_{\text{cum}} \) value slightly higher than P22, whereas its Rob value is lower, because \( x_3 \) and \( x_2 \) values are

<table>
<thead>
<tr>
<th>Solution</th>
<th>Weights</th>
<th>( x_1 )</th>
<th>( \hat{\mu}_1 )</th>
<th>( \hat{\mu}_2 )</th>
<th>( \hat{\sigma}_1 )</th>
<th>( \hat{\sigma}_2 )</th>
<th>Result</th>
<th>( B_{\text{cum}} )</th>
<th>QoP</th>
<th>Rob</th>
<th>Cum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pignatiello</td>
<td>P12</td>
<td>(0.01, 0.04, 0.30)</td>
<td>(−0.3866, 1.682, −0.672)</td>
<td>95.70, 57.94</td>
<td>1.04, 1.64</td>
<td>( E_{\text{Loss}} = 0.599 )</td>
<td>1.471</td>
<td>0.03</td>
<td>0.078</td>
<td>1.579</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P22</td>
<td>(0.01, 0.02, 0.80)</td>
<td>(−0.4058, 1.682, −0.8589)</td>
<td>95.07, 57.47</td>
<td>0.55, 1.43</td>
<td>( E_{\text{Loss}} = 1.403 )</td>
<td>1.189</td>
<td>0.03</td>
<td>0.064</td>
<td>1.283</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P32</td>
<td>(0.18, 0.01, 0.30)</td>
<td>(−0.8529, 1.682, −0.8067)</td>
<td>96.65, 55.98</td>
<td>0.00, 1.13</td>
<td>( E_{\text{Loss}} = 3.381 )</td>
<td>1.395</td>
<td>0.03</td>
<td>0.045</td>
<td>1.475</td>
<td></td>
</tr>
<tr>
<td>Ch’ng et al.</td>
<td>C12</td>
<td>(0.1, 0.7, 0.1, 0.1)</td>
<td>(−0.3652, 1.682, −0.9117)</td>
<td>94.60, 57.50</td>
<td>0.44, 1.39</td>
<td>D = 0.0296</td>
<td>1.183</td>
<td>0.03</td>
<td>0.061</td>
<td>1.274</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C22</td>
<td>(0.25, 0.25, 0.4, 0.1)</td>
<td>(−0.1307, 1.682, −0.9666)</td>
<td>94.05, 57.50</td>
<td>0.20, 1.33</td>
<td>D = 0.0563</td>
<td>1.184</td>
<td>0.03</td>
<td>0.057</td>
<td>1.271</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C32</td>
<td>(0.1, 0.1, 0.1, 0.7)</td>
<td>(−0.8820, 1.125, −1.2076)</td>
<td>87.87, 55.00</td>
<td>0.00, 0.53</td>
<td>D = 0.0923</td>
<td>2.425</td>
<td>0.03</td>
<td>0.021</td>
<td>2.476</td>
<td></td>
</tr>
<tr>
<td>Wu</td>
<td>W12</td>
<td>(0.01, 0.01, 0.3)</td>
<td>(−0.4720, 1.682, −0.9084)</td>
<td>95.16, 57.12</td>
<td>0.34, 1.33</td>
<td>D’ = 0.66</td>
<td>1.236</td>
<td>0.03</td>
<td>0.057</td>
<td>1.323</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W22</td>
<td>(0.01, 0.04, 0.8)</td>
<td>(−0.3051, 1.682, −1.0017)</td>
<td>93.20, 57.39</td>
<td>0.00, 1.21</td>
<td>D’ = 0.43</td>
<td>1.207</td>
<td>0.03</td>
<td>0.048</td>
<td>1.289</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W32</td>
<td>(0.01, 0.01, 0.5)</td>
<td>(−0.3800, 1.682, −1.0430)</td>
<td>93.86, 57.16</td>
<td>0.00, 1.20</td>
<td>D’ = 0.56</td>
<td>1.230</td>
<td>0.03</td>
<td>0.048</td>
<td>1.308</td>
<td></td>
</tr>
</tbody>
</table>
farther from the origin. This solution and others with the lowest Rob value are delimited by a square in Figs. 2 and 3.

The spread out of the estimated mean (\(\hat{\mu}_1\) and \(\hat{\mu}_2\)) and standard deviation (\(\hat{\sigma}_1\) and \(\hat{\sigma}_2\)) values for Pignatiello’s and Ch’ng et al.’s methods presented in Figs. 2–5 allows stating that they yielded distinct and equal or in close agreement solutions. An example of a distinct solution is C32 which has the lowest Rob value. However, C32 and similar ones with lower variance values (8 out of 424), which are not displayed in Fig. 2 because \(x_2 \neq 1.682\), have \(\hat{\mu}_1\) and \(\hat{\mu}_2\) too far from target, leading to \(B_{cum}\) and Cum values higher than those of all the other solutions. The solutions with the lowest \(B_{cum}\) and Cum values yielded by Ch’ng et al.’s method (C12 and C22) are delimited by a hexagon in Figs. 4 and 5, while those ones with lower Rob values displayed in these figures are delimited by a square.

As regards the results from Wu’s method, Table 3 shows that solution W22 gives \(B_{cum}\), Rob and Cum values at least in close agreement with W12 and W32. Its Cum and \(B_{cum}\) values are marginally less favorable than those of C12, C22 and P22. This is not surprising, because Wu’s method did not yield \(\hat{\mu}_2\) values on-target (see Fig. 6). On the other hand, W22, delimited by an ellipse in Figs. 6 and 7, gives a slightly lower Rob value than C12, C22 and P22.

4.3. Case 3 — low quality of predictions and unequal robustness

As in Case 1, some insignificant regressors are added to the model fitted to \(\hat{\mu}_1\). This makes the QoP fluctuate more severely when \(x_1\) is farther from the origin, whereas the robustness for the conversion (\(y_1\)) and thermal activity (\(y_2\)) responses improves. The estimated models are as follows:

\[
\hat{\mu}_1 = 81.8380 + 0.8234x_1 + 4.0717x_2 + 6.2127x_3 - 2.1774x_1^2 - 5.4906x_2^2 + 1.7698x_1x_2 + 11.4084x_1x_3 - 3.9496x_2x_3 + 2.8731x_1^3 - 1.6415x_2^3
\]

\[
\hat{\sigma}_1 = 4.5033 - 1.2540x_1 - 0.8972x_2 - 1.6415x_3^2
\]
\[
\hat{\sigma}_2 = 60.6935 + 3.5894x_1 + 0.3042x_2 + 2.1799x_3 + 0.5301x_1^2 - 0.1719x_2^2 - 0.2055x_1x_2 - 0.3381x_1x_3 + 0.0124x_1x_3 + 0.3714x_2x_3
\]

\[
\hat{\sigma}_1 = 2.0935 - 0.7300x_1^2 - 0.6744x_3^2
\]

The specification limits and targets for \( \hat{\sigma}_1 \) and \( \hat{\sigma}_2 \) are the same as in Case 2. The response weights and increments are equal to those used in Cases 1 and 2.

Fig. 9 shows that Ko et al.'s method did not yield solutions with Rob values as low as those of the other two desirability-based methods. In fact, Ko et al.'s method yielded solutions, displayed as dots in Fig. 9, with \( \hat{\sigma}_1 \geq 1.5 \) and \( \hat{\sigma}_2 \geq -1 \), whereas the desirability-based methods yielded solutions with \( \hat{\sigma}_1 < 1.5 \) and \( \hat{\sigma}_2 < 1 \). At first glance, this seems a serious drawback, although it does not mean that Ko et al.'s method yields compromise solutions with worse response properties than the other two methods. Indeed, QoP and Bcum have to be also taken into account in the evaluation of compromise solution.

In terms of Bcum values, Fig. 8 shows that Ko et al.'s method can yield solutions with high \( \hat{\mu}_1 \) values and \( \hat{\mu}_2 \) values on-target, which are in close agreement with those yielded by Ch'ng et al.'s method that are delimited by a rectangle. These solutions lead to Bcum values slightly lower than those yielded by Wu's method, because all their solutions have \( \hat{\mu}_2 \) off-target, such as Fig. 8 displays.

As regards the variance (QoP plus Rob), Table 4 shows that the solutions with the lowest variance values are C13, W13 and K13, respectively. However, these solutions are not the best ones in terms of Cum. The solutions C13 and W13 are even the worst due to higher Bcum values.

Looking at Cum values, there is no significant differences among the best solutions of the three methods. Nevertheless, Ch'ng et al.'s method yields two solutions (C23 and C33), which present different D values due to the different weights assigned to responses, with Bcum and Cum values slightly better than all the other solutions displayed in Table 4. The solution C23 also presents the lowest D values, in contrast with Ko et al.'s solution with the lowest expected loss (K23), which is not the best in terms of Cum value. The solution K23 is worse than K13 with respect to Bcum, QoP, Rob and Cum values, and worse than K33 with respect to Bcum, variance (QoP plus Rob) and Cum values.

Wu's method yielded two solutions with similar Cum values, W23 and W33, which have variance slightly higher than W13. These solutions are marginally worse than C23, C33 and K33 with respect to Cum due to higher Bcum and QoP values, although they present slightly better Rob values. In particular, W23 gives the lowest Rob value among the methods solutions with the lowest Cum value presented in Table 4, and the highest D value among the solutions yielded by Wu's method, displayed as squares in Figs. 8 and 9. Solutions W23 and W33 are delimited by an ellipse in these figures.

Like Cases 1–2, Wu's method yielded many (327 out of 480) solutions with \( x_2 \neq 1.682 \) (\( x_2 = 0 \)), leading to high Cum values (Cum = 4.756), whereas Ch'ng et al.'s and Ko et al.'s methods yielded 37 out of 424 with 1.51 < Cum < 2.44 and 24 out of 480 with 1.55 < Cum < 1.98, respectively. These solutions are not displayed in Figs. 8 and 9, because these figures only show solutions with \( x_2 = 1.682 \).

5. Discussion

To date, there is no effective procedure to guide the analyst in the selection of weights or priorities to responses. The optimization results can vary significantly if the weights change, and very little is usually known about how to choose their values. In practice, it is difficult, if at all possible, to determine beforehand the modifications required in the weights so as to produce a solution of interest, that is,
so as to know which response(s) will change and which is the direction and magnitude of that change.

To code the optimization methods usually used in the RSM framework, Matlab® is an alternative that avoids testing various weights to each response using the usual trial-and-error procedure, which may be a source of frustration and significant inefficiency, particularly when either the number of responses or control factors is large. Nevertheless, the range and increment for the weights assigned to responses are required. Their specification impacts on computational time, so those values must be specified taking into account this inconvenience. The analyst must be also aware that if they are not properly defined, satisfactory compromise solutions will be lost.

Regarding the results of Cases 1–3, there are no significant differences among Ch’ng et al.’s method and the others. All the methods yielded competitive compromise solutions for the three operating conditions tested here. Although the spread out of solutions may indicate that a particular method is more flexible in the sense that can yield solutions with either lower bias or variance values by changing the weights assigned to responses, the results show that a significant improvement in one property may degrade seriously other(s). This fact points out the intrinsic objective to multiresponse optimization problems: to identify a compromise solution among the generated responses. The methods evaluated here can do it, but Ch’ng et al.’s method may take some advantage over the others due to its major feature. This method requires a number of weights just equal to the number of responses, and to use it, the user only has to decide about the magnitude of the weights increment, because the sum of the weights assigned to responses must be equal to one. In practice, less than 1 min is enough for running the routine in Matlab® for the problems usually developed in the RSM framework and illustrated in this article. However, it is not guaranteed that the compromise solution with the lowest composite function value leads to responses with desired properties, namely low bias, high quality of predictions and high robustness if the weight increments are not appropriate. The same occurs whenever loss function-based methods are used and the loss coefficients are not properly specified. To specify loss coefficients taking into account different scales, relative variabilities, and relative costs is difficult [41, 43]. In practice, they may be not readily available or easily defined. Here, we tested loss coefficients around those used in Refs. [41] and [43]. Therefore, one cannot guarantee that the solutions presented here are the best ones that their methods can yield.

In the RSM framework few authors have addressed the evaluation of response’s properties to the extent it deserves. In general they focus on the objective function value used to select the compromise solution. However, that value depends on the weights assigned to responses and do not provide information on responses properties to the analyst. What is known is that a higher or a lower value is desirable, depending on the optimization scheme used, which does not guarantee that desired responses properties are achieved. Such as [31], the function values are not similar response properties, however they have different expected values. For example: C11 and C12 have similar expected values, however their desirability values are not the same. Furthermore, V31 and V51 have similar response properties, however they have different expected loss values. Hence, the analyst may take as reference the Cum value of response(s) of interest and selecting solutions feasible in practice.

The measures proposed in Ref. [40] were used in this article and the results of Cases 1–3 show that solutions with Bcum, QoP and Rob values similar or slightly better than those obtained through the objective functions can be achieved. To select the compromise solutions the analyst may take as reference the Cum value (cumulative value of Bcum, QoP and Rob). This measure is more useful than the loss and desirability composite functions, because they depend on the weights assigned to responses and give different results for similar solutions. For example: C11 is similar to C41, however their desirability values (D values) are very different; V31 and V51 have similar response properties, however they have different expected loss values; P12 has response properties worse than P32, however its expected loss is lower; K23 and V33 have similar response properties, however they have different expected loss values.

These results provide evidence that the optimization measures are effective, but this does not mean that they are the panacea to achieve optimal solutions. They serve to assess response’s properties but not...
for generate compromise solutions. Moreover, the analyst must be aware that points in non-convex response surfaces cannot be captured by weighted sums like those of the objective functions tested here [45], even if the proposed measures are used.

6. Conclusions

This article provides a review on desirability-based methods and evaluates, under adverse variance conditions, two of the less sophisticated ones, namely the methods proposed by Ch’ng et al. [27] and Wu [28]. To assess the properties of the compromise solutions at variables setting with respect to bias, quality of predictions and robustness, optimization measures were used. These measures can be used along with any method of practitioner’s interest, and have proved to be useful for selecting a compromise solution taking into account the responses properties of the analyst or decision-maker preference.

The results show that none of the methods outperforms the others. Nevertheless, Ch’ng et al.’s method presents appealing features and along with the presented measures allows the analyst to identify compromise solutions with response properties similar or slightly better than those ones yielded by other methods used in practice. This gives confidence to suggest this approach for solving real life problems developed in the RSM framework as alternative to more sophisticated methods or approaches. This does not mean that the problem of choosing a method for optimizing multiple mean and standard deviation responses is resolved. A large number of methods, including parametric and non-parametric methods, from those requiring little information to the methods that require extensive information on each objective, have been reported in the literature. To compare their performance in adverse variance conditions and assess their ability to capture points in non-convex response surfaces are issues that deserve further research.

Acknowledgment

The authors are grateful to reviewers for their comments and suggestions. The Fundação para a Ciência e Tecnologia (FCT) through the PIDDAC Program funds (INESC-ID multiannual funding) supported the second author.

References