Multi-response optimization design based on Non-parametric error-corrected method

GAO Tingyu¹

¹ Master's Degree, School of Economics and Management, Nanjing University of Science and Technology, China

E-mail:1721922493@qq.com

Abstract. The construction of a response surface model is critical to the experimental results. Traditional model construction method is parametric method. The parametric estimates may be highly biased, and the optimal control factor settings can be miscalculated if the models are not correctly specified. To solve these problems, this paper proposes a new multi-response optimization design method, Non-parametric error-corrected method. The nonparametric method provides a very useful alternative when researchers don’t have any information about the form of underlying functions. Finally, the hybrid genetic algorithm is used to achieve global optimization aiming at the expected quality loss function the validity of the method was verified by the experimental data of the femtosecond laser micro/nano-machining.

1. Introduction
Response surface method (RSM) is generally considered to be a statistical analysis method to obtain optimal process parameters, the method which contains experimental design, response model and optimization methods [1]. With the diversification of customer demand levels, it is often necessary to consider multiple quality characteristics. The optimization of multi-response surfaces is increasingly important, this paper will focus on the optimization of multi-response parameters.

Multi-response surface optimization design usually includes three stages which contain experimental design, model building and parameter optimization. The most important is model building. However, as the famous statistician Box [2] said, "All models are wrong, but some models are useful." Traditional response surface modeling methods often assume the structure of response surface models is sure (such as dual response surface methods, generalized linear models, etc.). However, in many cases, these overly perfect assumptions often lead to unreliable research results and even false conclusion. Therefore, the construction of response model structure needs to make full use of known data information rather than perfect assumptions.

For multi-response optimization problems, Wang et al. [3] used SMR and SUR models combined with multiple quality loss functions and posterior probability methods under the Bayesian statistical modeling framework. The method comprehensively consider the correlation between multiple responses and robustness [4, 5]. Jun et al.[6] pointed out that "traditional parametric regression methods usually need to assume the model structure between the response and the factor before fitting the response surface model, which often does not adapt to the actual situation. Or even wrong research conclusions. " Therefore, in the multi-response optimization design, some researchers try to use non-parametric to develop response surface modeling and obtain a series of research results. For example, the general parameter method can cause bias in the mean estimation, or lead to higher variance.
Pickle[7] also studied the mean square error of the mean and variance models under the non-parametric methods in different situations to prove the effectiveness of this method. Wan[8] believes that non-parametric methods can capture data information that is ignored in parametric methods. Wan and Birch [9] also show that the nonparametric method is better than traditional parametric approach when the problem is more difficult.

In this paper, a new multi-response optimization method is proposed that combines a non-parametric approach with Bayesian sampling techniques.

### 2. Nonparametric response surface modeling

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

The non-parametric method is based on data driving, and the form of the regression function and the random error distribution are not strictly required.

Once the model function form of the parameter method cannot be accurately expressed, the response surface modeling can be performed using the non-parametric method. The non-parametric does not need to set the basic form of the model function in advance. Non-parametric methods include kernel function regression, local polynomial regression, and spline regression. Local polynomial regression is essentially an extension of kernel function regression. Both belong to the family of local polynomial regression. The kernel function is a weight function.

#### 2.1. Kernel function

The non-parametric method uses local linear regression (LLR), which is a first-order polynomial regression. The key part is the processing of the kernel function. The shape and value range of the kernel function can reflect the weight of the amount of data used at the estimated point \( x_0 \) of the response estimate \( f(x_0) \). At the point \( x_0 = (x_{01}, x_{02}, \ldots, x_{0d}) \) that needs to be predicted, we define the kernel function as:

\[
K \left(x_0, \tilde{x}_i \right) = \frac{1}{b^d} \prod_{j=1}^{k} K \left( \frac{x_{0j} - \tilde{x}_j}{b} \right)
\]  

(1)

In equation (1), \( \tilde{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id}) \) refers to any arbitrary point, \( K \left( \frac{x_{0j} - \tilde{x}_j}{b} \right) \) is a univariate kernel function, \( i \) refers to subscript values at multiple design points, \( x_0 \) is a prediction point, \( b \) and is a bandwidth. When non-parametric methods are used for the mean or variance models, because the control factors that affect these two different models may be different, generally a second-order model is used for the mean model and a first-order model for the variance model, so different kernel functions will be used. There are many common kernel functions, such as Gaussian function, uniform function and trigonometric function. However, according to the research of Simonoff [10], the change of kernel function does not have a great influence on the estimated value. This article will use the most popular Gaussian kernel function in the previous literature, which has the form: \( K (u) = e^{-u^2} \).

#### 2.2. Bandwidth choose function

The smoothness of the estimated function \( K \left(x_0, \tilde{x}_i \right) \) depends on the bandwidth parameter \( b \). Mays et al.[11] proposed a penalized cross-validation technique, \( PRESS^* \), to select the appropriate bandwidth \( b \) by minimizing the \( PRESS^* \).

\[
PRESS^* = \frac{PRESS}{n - trace \left(H^{LLR} \right) + \left(n - (k + 1) \right) \left( \frac{SSE_{max} - SSE_b}{SSE_{max}} \right)}
\]  

(2)
In formula (2), $SSE_{\text{max}}$ represents the largest error sum of squares over all possible bandwidth values; $SSE_b$ refers to the error sum of squares at a certain bandwidth value $b$; $k$ is the number of regressors. The prediction error sum of squares, $PRESS$, is given by $PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2$, where $\hat{y}_{i,-i}$ refers to the estimated response obtained by leaving out the $i^{th}$ observation when estimating at location $x_i$. $H^{LLE}$ is the smooth matrix obtained by $LLE$. $h_i^{LLE}$ performs non-parametric processing at each design point.

$$H^{LLE} = \begin{bmatrix} h_1^{LLE} \\ h_2^{LLE} \\ \vdots \\ h_n^{LLE} \end{bmatrix}$$

(3)

3. The proposed method

The multi-response surface model based on the non-parametric method can generally be assumed as:

$$\hat{y}_i = f(x_i) + g^{1/2}(x_i^*; \gamma) \epsilon_i$$

(4)

$h(x_i, \beta)$ represents the parameter method part, $f(x_i)$ represents the non-parameter part, $g^{1/2}(x_i^*; \gamma)$ is the transformation of the variance model, and $\epsilon_i$ is the random error. However, the above approach to random error assumptions obeying the standard normal distribution may not be consistent with the actual situation and cannot realistically reflect the distribution characteristics of the model’s random error; in addition.

Drawing on the idea of reference [9] which uses the Bayesian sampling method to obtain the random error term. This article will modify the random error term in formula (4), where the random error term is replaced by $W \sqrt{v/U}$. The error corrected response surface model using Bayesian sampling techniques is:

$$\hat{y} = \hat{y}^{LLE} + W \sqrt{v/U}$$

(5)

Where $v$ is the degree of freedom, $N$ is the sample size, and $W$ is a multivariate normal random variable with zero mean vector and variance covariance matrix $H^{-1}$. $z(x_i)$ is the observation value of $q \times 1$ at the observation point, where $X$ is the second-order factor model matrix and $Y$ is the $N \times p$ response matrix. $\hat{B}$ is the fitting coefficient vector, $Z$ is a $q \times N$ matrix formed by $z(x_i)$ vectors of number $N$. $U$ is a chi-square random variable whose degree of freedom $v$ does not depend on $W$. Under the framework of the above response surface model, combined with the dual response surface modeling idea[12], the response mean and variance models were constructed. Then use the multivariate quality loss function to obtain the desired objective function as follows:

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

(6)
The above formula (6) is essentially a new model constructed using the idea of combinatorial modeling. The model combines the advantages of non-parametric methods and Bayesian sampling methods without assuming a parametric model structure. Therefore, compared with previous research methods, the use of error-corrected non-parametric response surface models can better extract the characteristics of the experimental data and build more accurate and realistic response surface models.

To address the problems in the optimization design of multi-response surfaces, this paper proposes a new modeling technique combining non-parametric methods and Bayesian sampling techniques, with the following implementation steps figure 1.

4. Case Analysis
This example comes from the literature [13], and the purpose of this experiment is to realize the micro-hole fabrication process of a satellite chip through the Femtosecond Laser Micro-Nano Machining Center of the Department of Mechanical and Manufacturing Engineering, University of Calgary, Canada. In this experiment, the controllable factors were: Average power $x_1$ (mw), pulse frequency $x_2$ (Hz) and cutting speed $x_3$ (mm/s). The output response of the test is diameter $y_1$ and roundness $y_2$. In addition, roundness is related to the area of the micropores and the spindle, which is specifically defined by the formula.

$$\text{Roundness} = 4\pi \times \frac{\text{Area}}{\text{Spindle}^2}$$  \hspace{1cm} (7)

The purpose of this experiment was to determine the optimal setting of the three variables so that the target value of the diameter $y_1$ is maintained at the level of 21 and the roundness $y_2$ is maximized. Response roundness $y_2$, although it is a desired quality characteristic, the target value can be viewed as 1 according to Equation (7). Among them, laser micro-drilling experimental data were obtained by central composite design (CCD) as shown in table 1.

Table 1. The experimental design results of the femtosecond laser micro/nano-machining.

| Test order | Control factor | Responses |
|-----------|----------------|-----------|
|           | $x_1$ | $x_2$ | $x_3$ | $y_1$ | $y_2$ |
| 1         | -1   | -1   | -1   | 16.43 | 0.965 |
| 2         | 1    | -1   | -1   | 23.8  | 0.94  |
| 3         | -1   | 1    | -1   | 15    | 0.968 |
| 4         | 1    | 1    | -1   | 16.53 | 0.923 |
| 5         | -1   | -1   | 1    | 14.47 | 0.922 |
| 6         | 1    | -1   | 1    | 22.81 | 0.924 |
| 7         | -1   | 1    | 1    | 14.04 | 0.921 |
| 8         | 1    | 1    | 1    | 17.43 | 0.929 |
Throughout the analysis of the experiment, it was assumed that the vectors constituted by the factor effects in the regression model were as follow:

\[
z(x) = \left( 1, x_1, x_2, x_3, x_1 x_2, x_2 x_3, x_1 x_3, x_1^2, x_2^2, x_3^2 \right)
\] (8)

According to the optimization function \( PRESS^{**} \) of Equation (2). The figure 2 and 3 display the \( PRESS^{**} \) plots. The minimum value of the function \( PRESS^{**} \) denotes the optimal bandwidth \( b \). The optimal bandwidth value \( b \) for both responses is 0.58 after calculation.

**Figure 2.** Function \( PRESS^{**} \) image of \( y_1 \).
Figure 3. Function $PRESS^{**}$ image of $y_2$.

Optimize parameters with the help of a hybrid genetic algorithm as an objective function of equation (6). The parameter optimization results are $x_1 = 0.8196$, $x_2 = -0.1282$, $x_3 = -0.5322$. The expected quality loss value is 2.0941.

5. Conclusion
In this paper, a multi-response surface model is constructed by effectively combining non-parametric methods and Bayesian sampling techniques. The appropriate bandwidth value is selected by function $f_1$, and then the objective function is constructed using the non-parametric-error-correction model of this paper in conjunction with the expected quality-loss function and optimized by the hybrid genetic algorithm parameters. The validity of the methodology has been demonstrated through cases, but the methodology is still somewhat crude in terms of error handling, and future research in this area could be continued. It is also possible to refine the whole study by considering the reliability test of the results. In conclusion, the current research is still relatively shallow, and there are still many directions for refining the research, which will be strengthened in the future to make more mature findings.

References
[1] He Z, Zhu P F and Park S H 2012 A robust desirability function method for multi-response surface optimization considering model uncertainty European Journal of Operational Research 221(1) 241-7
[2] Brown R A, Box G E P and Draper N R 1987 Empirical model-building and response surfaces New York: John Wiley & Sons
[3] Wang Jianjun, Ma Yizhong, Ouyang Linhan, Sun Jinsheng and Liu Jian 2016 Bayesian modeling and optimization of multi-response robust parameter design Journal of Management Sciences in China 19(02) 85-94
[4] Wang J, Ma Y, Ouyang L and Tu Y 2020 Bayesian modeling and optimization for multi-response surfaces Computers & Industrial Engineering 142
[5] Wang J, Ma Y, Ouyang L and Tu Y 2016 A new Bayesian approach to multi-response surface optimization integrating loss function with posterior probability European Journal of Operational Research 249(1) 231-7
[6] Jun Young Min 2005 Off-line quality control by robust parameter design B S Iowa State University
[7] Pickle S M, Robinson T J, Birch J B and Anderson Cook C M 2008 A approach to robust parameter design Journal of Statistical Planning and Inference 138(1) 114-31
[8] Wen Wan 2007 Techniques for Multi-Response Optimization Virginia Polytechnic Institute & State University

[9] Wan W and Birch J B 2011 A semiparametric technique for the multi-response optimization problem Quality and Reliability Engineering International 27(1) 47-59

[10] Simonoff J S 1996 Smoothing Methods in Statistics Springer Series in Statistics New York

[11] Mays J E, Birch J B and Alden Starnes B 2001 Model robust regression: combining parametric nonparametric and semiparametric methods Journal of Nonparametric Statistics 13(2) 245-77

[12] Wang Jianjun and Ma Yizhong 2012 Dual response surface methodology based on generalized linear models and application on robust design Systems Engineering and Electronics 34(11) 2306-11

[13] Wang Jianjun and Tu Yanan 2018 Optimal design for multiple responses considering predicted response variability Systems Engineering and Electronics 40(08) 1794-802