

# 结合半参数方法和贝叶斯抽样技术的 多响应优化设计

汪建均, 郜婷玉, 杨世娟

(南京理工大学经济管理学院, 江苏 南京 210094)

**摘要:** 响应曲面模型的构建对多响应优化设计结果的影响至关重要. 传统的响应曲面模型会事先对模型结构做出一系列的假设. 然而, 在面向复杂产品的质量设计时往往会出现模型结构错误设定的情况. 结合半参数方法和贝叶斯抽样技术提出了一种新的多响应优化设计方法, 以解决多目标之间的冲突、模型结构的不确定以及预测响应值的波动性问题. 首先, 利用半参数方法建立可控因子与响应之间的响应曲面模型; 其次, 基于贝叶斯抽样技术对模型误差进行修正, 建立基于误差修正的响应曲面模型, 并在此基础上构建多目标优化函数; 然后, 利用混合遗传算法进行寻优, 获得最佳的参数设计值. 另外, 基于贝叶斯抽样技术对优化结果进行稳健性评估, 以考察优化结果的可靠性. 最后, 通过一个案例分析说明所提方法的有效性. 结果表明, 所提方法能有效地解决模型结构不确定以及小样本数据对研究结果的影响, 获得更加稳健可靠的优化结果.

**关键词:** 多响应优化设计; 半参数方法; 质量损失函数; 响应曲面方法; 贝叶斯方法

中图分类号: O212.6

文献标识码: A

文章编号: 1000-5781(2021)05-0697-12

doi: 10.13383/j.cnki.jse.2021.05.010

## Multi-response optimization design based on semi-parametric methods and Bayesian sampling techniques

Wang Jianjun, Gao Tingyu, Yang Shijuan

(School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** Constructing an accurate response surface model is critical to achieving reliable experimental results. Traditional methods tend to make a series of assumptions about the model structure before model fitting. However, misspecification of the model structure often occurs in the quality design of complex products. This paper proposes a new multi-response optimal design method to address the conflicts between multi-objectives, the model structure uncertainty, and the volatility of the response. Firstly, a semi-parametric method is used to establish a response surface model. Secondly, the model error is corrected based on Bayesian sampling technology. Thirdly, a multi-objective optimization function is constructed, and a hybrid genetic algorithm is used to obtain the optimal setting. Finally, a case study is used to illustrate the effectiveness of the proposed method. The results show that the proposed method can effectively solve the model structure uncertainty and the influence of small samples.

**Key words:** multi-response optimization; semi-parametric methods; quality loss function; response surface methods; Bayesian methods

## 1 引言

响应面方法(response surface methodology, RSM)通常被认为是利用试验设计、响应模型和优化方法获取最优工艺参数的一种统计分析方法<sup>[1]</sup>. 统计学家 Myers 等<sup>[2]</sup>曾明确指出“从更广泛的意义上看, 响应曲面方法已经成为整个工业试验的核心”. 随着顾客需求层次的多样化以及产品性能的稳步提升, 在实际的产品与工艺过程中, 往往需要考虑多个质量特性, 因此多响应曲面优化设计问题在持续性质量改进活动中显示出越来越重要的作用<sup>[3-5]</sup>.

多响应曲面优化设计通常包括试验设计、模型构建以及参数优化三个阶段, 其中响应曲面模型构建至关重要<sup>[6]</sup>. 在多响应曲面优化设计中, 研究者首先需要构建合适的度量指标. 然而, 由于涉及到多个质量特性, 因此在度量指标构建中往往需要考虑多响应之间的相关性、模型参数的不确定性、预测响应的波动以及过程的稳健性<sup>[7]</sup>. 在多响应曲面的优化设计中, 常见的优化策略是通过数据降维方法将多响应的优化问题转换为单一指标的优化问题, 其中具有代表性的方法为: 满意度函数方法、多元质量损失函数和后验概率方法等. 比较而言, 满意度函数方法简单实用, 并且能够在一些常用统计软件如 MINITAB 中加以使用, 因此引起一些研究者和质量工程师的广泛关注. 例如, He 等<sup>[8]</sup>综合考虑稳健性和最优性提出了一个新的满意度函数模型, 有效地解决了传统满意度函数未考虑多响应优化过程的稳健性问题. 然而, 满意度函数方法难以有效地考虑多响应之间的相关性、预测响应的波动以及模型参数的不确定性. 为此, 一些研究者运用质量损失函数来实现多响应优化设计. 例如, Ko<sup>[9]</sup>扩展传统多元质量损失函数, 进一步地考虑多响应的预测性能和稳健性能, 有效地兼顾过程的偏差、稳健性以及预测性能之间的平衡. 然而, 正如文献<sup>[9]</sup>中所指出的那样“所提出的多元质量损失函数未能考虑模型参数不确定性对优化结果的影响”.

在响应曲面建模过程中, 研究者期望通过响应曲面模型来反映试验因子与多响应之间精确的函数关系, 从而获得更为合理的参数优化结果. 但是, 正如著名统计学家 Box<sup>[10]</sup>所言“所有的模型都是错误的, 但是有些模型是有用的”. 传统的响应曲面建模方法通常事先假定响应曲面模型结构, 然后采用参数方法加以估计. 然而, 在很多情况下, 所假设的响应曲面模型结构并不符合客观的实际情况, 从而导致无法获得可靠的研究结果, 甚至出现错误的研究结论. 因此, 在响应曲面模型构建过程中需要考虑模型不确定性(模型参数以及模型结构不确定)对优化结果的影响. 为此, Peterson<sup>[11]</sup>在标准的多响应曲面(standard multi-response models, SMR)框架下利用贝叶斯方法考虑了模型参数不确定性对优化结果的影响, 进而利用蒙特卡洛方法分析了优化结果的可靠性, 即未来响应预测值落在产品规格限内的概率. Peterson 等<sup>[12]</sup>指出 SMR 模型通常假设各个响应的模型结构是相同的, 即模型所包含的因子效应是完全相同的. 然而, 在很多情形下响应模型所包含的因子效应往往是完全不同的. 鉴于此, Peterson 等<sup>[12]</sup>在似不相关回归模型(seemingly unrelated regression, SUR)的框架下利用贝叶斯方法开展了多响应优化设计. 针对多响应优化问题, 汪建均等<sup>[13]</sup>在贝叶斯统计建模框架下利用 SMR 和 SUR 模型结合多元质量损失函数和后验概率方法, 全面地考虑多响应之间的相关性、多元过程的稳健性、多目标之间的冲突以及优化结果的可靠性. 上述响应曲面建模方法均事先对模型结构和响应分布等做出一些假设. 然而, Min 等<sup>[14]</sup>指出“传统的参数回归方法通常需要在拟合响应曲面模型之前假设响应与因子之间的模型结构, 这往往与实际情况并不相符, 从而导致出现不可靠的研究结果, 甚至是错误的研究结论”. 比较而言, 非参数方法不需要事先假定模型结构, 而是利用试验数据本身所包含的信息自适应地拟合响应曲面模型, 往往更加符合实际情况. 因此, 在多响应优化设计中, 一些研究者尝试地利用非参数方法开展响应建模工作, 并取得一系列的研究成果. Vining 等<sup>[15]</sup>认为参数模型不能较好的刻画方差模型与各因子间的关系. 因此他们采用基于核估计的非参数方法构建均值模型与方差模型, 并指出非参数方法所构建的响应曲面模型的预测性能要优于基于二阶多项式模型所构建的响应曲面模型. Ouyang 等<sup>[16]</sup>结合 Bootstrap 方法和 Pareto 优化提出一种集成的径向基函数(radial basis function, RBF)神经网络方法, 并利用该非参数方法进行了参数优化设计. 另外, 一些研究者也考虑利用非参数方法对非正态响应的质量设计

问题进行了研究. 例如, Fox 等<sup>[17]</sup>利用非参数方法对非正态响应的试验设计进行了分析. 针对具有重尾分布的非正态响应数据, Lim 等<sup>[18]</sup>结合平滑样条方法提出了一种基于非参数分位数回归的置信区间新方法, 有效地解决上述非正态响应的质量设计问题.

虽然非参数方法不需要对模型结构进行事先的假设, 从而在建模方面具有良好的灵活性, 能够直接根据观测数据拟合响应的均值和方差模型. 但是非参数方法往往依赖于相对大的样本或空间填充设计, 样本量偏小会可能导致响应波动较大. 针对非参数方法的不足之处, 一些研究者提出了一些能够处理小样本试验数据的半参数方法. 因此, 基于半参数的多响应优化设计方法也引起一些研究者的广泛关注和兴趣. Wan<sup>[19]</sup>认为非参数方法对稀疏数据进行响应曲面建模时, 通常会导致其响应方差过大, 并指出半参数方法则能较好地处理该问题. Pickle<sup>[20]</sup>通过比较三种方法(即参数方法、非参数方法和半参数方法)的均值模型和方差模型的均方误差值, 验证了所提半参数方法的有效性. 他们提出了两种半参数的建模策略, 一种是模型稳健回归方法1(model robust regression 1, MRR1); 另一种是模型稳健回归方法2(model robust regression 2, MRR2). 在 MRR1 中的组合成分中, 一部分是响应的参数拟合项, 另一部分是响应的非参数拟合项. 该方法将这两项的权重和设为1, 只要假设其中某一项为未知的混合参数即可. 在 MRR2 的组合成分中, 一部分为响应的参数拟合项, 另一部分则替换为非参数残差拟合. 因此 MRR2 默认参数项的权重为 1, 非参数项为小于 1 的权重系数(即混合参数). 其中, 混合参数的选择是基于数据驱动的渐近最优值表达式. Mays 等<sup>[21,22]</sup>通过仿真方法来选择非参数模型的带宽和混合参数, 从而获得更为稳健的半参数估计结果. 此外, 在该文献中, 他们还给出了一种新的惩罚准则作为选择策略, 并阐述了单独使用参数或者非参数方法的局限性. Mays 等<sup>[23]</sup>对混合参数进行渐近最优估计和模型可靠性估计, 并通过一个小型仿真实验验证了所提方法的有效性. Wan 等<sup>[24]</sup>提出了以期望函数为优化目标函数的多响应稳健回归方法, 他们结合试验数据和仿真数据对参数方法和半参数方法进行了对比研究. 研究结果表明: 与参数方法相比, 基于半参数优化方法的偏差和方差都更低.

目前, 基于半参数方法的多响应优化设计仍然存在一些问题没有得到很好地解决. 例如, 在多响应优化设计中, 一些研究者往往没有考虑模型不确定性以及随机误差所引起的预测响应波动, 从而导致获得不可靠甚至错误的研究结论. 另外, 由于多响应目标之间的冲突以及模型结构等不确定因素影响, 其优化结果的可靠性(预测响应值落在所给定的规格限内的概率)也有待进一步研究. 针对上述问题, 本文结合半参数方法和贝叶斯抽样技术提出一种新的多响应优化方法. 首先, 通过半参数方法建立可控因子与响应之间的多响应面模型; 其次, 利用质量损失函数建立期望目标函数, 同时通过贝叶斯抽样方法构建随机误差修正模型; 然后, 通过混合遗传算法对所构建的模型进行参数优化, 获得最佳的参数设计值. 最后, 结合贝叶斯抽样技术对所获得参数优化结果进行稳健性评估, 以考察其优化结果的可靠性.

## 2 非参数和半参数建模方法

当模型结构错误设定时, 传统的参数回归方法可能会造成试验数据拟合不合理. 尤其是当研究问题比较复杂、模型不确定性较高时, 这种模型拟合不恰当问题就更加突出. 因此 Anderson-Cook 等<sup>[25]</sup>提出了采用局部回归思想的非参数方法对试验设计进行拟合, 以应对较为复杂的稳健参数设计问题. 在通常情况下, 在无法确定输入因子与输出响应之间的函数关系时, 分析人员可以采用数据驱动方法来局部回归拟合响应曲面模型. 由于数据驱动方法不仅可以拟合线性模型也可以拟合高度复杂的非线性模型, 因此基于数据驱动的非参数或半参数方法可以较好地构建稳健的响应曲面模型, 从而能够处理高度复杂的响应曲面优化设计问题.

### 2.1 非参数响应曲面建模

非参数方法是基于数据驱动的响应曲面建模方法, 其回归函数的形式和随机误差分布都不作严格要求. 针对某一点的估计值, 那些接近该点的响应通常假设比那些远离该点的响应包含更多的信息. 因此, 为了得

到一个光滑的函数,一些非参数方法会使用局部加权平均原理.局部平均的基本思想等同于寻找局部加权最小二乘估计量的过程.若无法确定响应曲面模型的具体形式,分析人员通常会考虑利用非参数方法进行响应面建模.非参数无需事先设定模型函数的基本形式,因此更加符合实际情况.常见的非参数方法有核函数回归、局部多项式回归和样条回归等方法.其中,局部多项式回归本质上是核函数回归的一种拓展.

非参数方法通常使用局部线性回归(local linear regression, LLR),其中 LLR 是 1 阶多项式回归.非参数估计关键在于如何确定核函数.核函数的形状和值域可以反映响应估计值  $f(\mathbf{x}_0)$  在估计点  $\mathbf{x}_0$  处所用数据量的权重.由于核函数方法不利用有关数据分布的先验知识,对数据分布也不附加任何假定,是一种从数据本身出发研究数据分布特征的方法.

针对需要预测的点  $\mathbf{x}_0 = (x_{01}, x_{02}, \dots, x_{0k})$ ,核函数可以定义为

$$K(\mathbf{x}_0, \tilde{\mathbf{x}}_i) = \frac{1}{b^k} \prod_{j=1}^k \kappa\left(\frac{x_{0j} - \tilde{x}_{ij}}{b}\right), \quad (1)$$

其中  $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ ,指任意某一点,  $\kappa((x_{0j} - \tilde{x}_{ij})/b)$  是单变量核函数,  $i$  是指多个设计点处的下标值,  $\mathbf{x}_0$  是预测点,  $b$  是带宽.

当对均值或者方差模型都使用非参数方法时,由于影响这两种不同模型的控制因子可能不同,一般对均值模型使用二阶模型,对方差模型使用一阶模型,所以将使用不同的核函数.常见的核函数为高斯函数、均匀函数和三角函数等.根据 Simonoff<sup>[26]</sup>的研究,核函数的变化并不会对估计值有很大的影响,所以本文将使用以往文献中最为流行的高斯核函数,其形式为  $\kappa(u) = e^{-u^2}$ .  $K(\mathbf{x}_0, \tilde{\mathbf{x}}_i)$  中估计函数的平滑性取决于带宽参数  $b$ . Mays 等<sup>[21]</sup>提出了新的惩罚预测误差平方和技术(penalized prediction error sum of squares, PRESS\*\*)进行合适的带宽选择,通过最小化 PRESS\*\* 的函数值来寻求带宽  $b$  的最优值.这里

$$\text{PRESS}^{**} = \frac{\text{PRESS}}{n - \text{tr}(\mathbf{H}^{\text{LLR}}) + (n - (k + 1))(\text{SSE}_{\text{max}} - \text{SSE}_b) / \text{SSE}_{\text{max}}}, \quad (2)$$

其中  $\text{SSE}_{\text{max}}$  代表所有可能带宽值下的最大误差平方和;  $\text{SSE}_b$  是指特定某个带宽值下的误差平方和;  $k$  是回归因子的个数,一般是控制变量的个数;分子部分是预测误差平方和(prediction error sum of squares, PRESS).

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_{i,-i})^2, \quad (3)$$

其中  $\hat{y}_{i,-i}$  是指通过留一法在点  $\mathbf{x}_i$  处的估计值.

留一法是指交叉验证方法的一种,假设样本数据集中有  $N$  个样本.将每个样本单独作为测试集,其余  $(N - 1)$  个样本作为训练集,这样得到了  $N$  个分类器或模型,用这  $N$  个分类器或模型分类准确率的平均数作为此分类器的性能指标.每一个分类器或模型都是用几乎所有的样本来训练模型,最接近样本,这样评估得出的结果与训练整个测试集的期望值最为接近;  $\mathbf{H}^{\text{LLR}}$  是非参数方法 LLR 得到的平滑矩阵

$$\mathbf{H}^{\text{LLR}} = \left( (\mathbf{h}_1^{\text{LLR}})^{\text{T}}, (\mathbf{h}_2^{\text{LLR}})^{\text{T}}, \dots, (\mathbf{h}_n^{\text{LLR}})^{\text{T}} \right)^{\text{T}}, \quad (4)$$

其中  $(\mathbf{h}_i^{\text{LLR}})^{\text{T}}$  是对设计点  $i$  进行非参数处理得到的均值模型中因子变量矩阵的一个系数矩阵,  $i = 1, 2, \dots, n$ .

非参数方法具有良好的灵活性,能够直接根据观测数据拟合均值响应模型,但是这些拟合模型往往具有较高的方差.非参数方法往往依赖于相对大的样本或空间填充设计.与非参数方法不同的是,半参数方法并不要求大量的样本.在此,将重点介绍两种常见的半参数方法.

## 2.2 半参数响应曲面建模

半参数方法顾名思义,就是结合传统的参数方法和基于数据驱动的非参数方法的一个组合模型方法.针对以往的传统参数方法无法处理:第一,对基础模型只有部分了解;第二,数据本身波动较高.这种情况下,

用参数方法进行模型估计其误差会较大. 虽然非参数方法可以较好地处理上述问题, 但是非参数方法对试验数据的样本量有较高的要求. 半参数方法则可以充分发挥二者的优势, 较好地解决模型结构不确定和试验数据样本量小的问题. 参数方法通常使用经典的最小二乘法 OLS 进行参数估计, 非参数方法则采用局部线性回归 LLR 进行估计, 因此通过将传统的参数估计和基于数据驱动的非参数方法进行有机整合就可以构成半参数方法. 在此, 将介绍两种经典的半参数方法.

### 2.2.1 MRR1 方法

Robinson 等<sup>[27]</sup>提出了一种考虑均值响应建模的半参数方法, 即通过一个混合参数  $\lambda$  结合参数和非参数部分. 则可构建混合模型

$$\hat{\mathbf{y}}^{\text{MRR1}} = \lambda \hat{\mathbf{y}}^{\text{LLR}} + (1 - \lambda) \hat{\mathbf{y}}^{\text{OLS}}, \quad (5)$$

其中第一部分是通过对非参数回归得到的拟合向量和混合参数  $\lambda$  的乘积, 第二部分是参数回归得到的拟合向量和剩余比例部分的乘积. MRR1 通过给参数部分和非参数部分分配比例系数得到半参数模型, 但是这种组合模型的波动在以往文献中被证实较高, 其中  $\lambda \in [0, 1]$  而这个  $\lambda$  的选择法, 类似于非参数部分的带宽  $b$  的选择, 本质上是一种偏差方差的平衡, 混合参数  $\lambda$  的渐近最优值用下面的数据驱动表达式来求得, 即

$$\hat{\lambda}_{\text{opt}}^{\text{MRR1}} = \langle \hat{\mathbf{y}}_{-i}^{\text{LLR}} - \hat{\mathbf{y}}_{-i}^{\text{OLS}}, \mathbf{y} - \hat{\mathbf{y}}^{\text{OLS}} \rangle / \|\hat{\mathbf{y}}^{\text{LLR}} - \hat{\mathbf{y}}^{\text{OLS}}\|^2,$$

其中  $\hat{\mathbf{y}}_{-i}^{\text{OLS}}$ ,  $\hat{\mathbf{y}}_{-i}^{\text{LLR}}$  分别指的是该位置的留一法最小二乘法拟合值和留一法局部线性回归法拟合值.  $\langle \cdot, \cdot \rangle$  是指内积法,  $\|\cdot\|$  指标准欧氏范数.

该半参数方法可以通过平滑估计获取数据中的异常信息. 此外, MRR1 方法不仅可以弥补参数方法整体拟合偏差大的问题, 而且能够处理非参数方法过度拟合的局限性, 从而带来响应面设计的稳健性. 如果参数部分估计和非参数部分估计同时较高或者较低, MRR1 估计也会过高或者过低. 因此, 有必要引入 MRR2 半参数方法以纠正上述错误.

### 2.2.2 MRR2 方法

MRR2 方法是 Mays 等<sup>[22]</sup>通过改进 MRR1 方法获得一种新的半参数方法. 与 MRR1 方法比较而言, MRR2 方法在建模组合策略上有所不同. 第一, MRR2 方法将对参数方法得到的残差进行非参数处理, 得到非参数残差部分. 然后, 将非参数残差部分和参数拟合部分进行组合. 第二, MRR1 将非参数和参数部分各自设定未知的权重系数, 权重和为单位 1. 而 MRR2 方法则将参数拟合部分权重赋值为 1, 非参数残差部分权重用混合参数  $\lambda$  ( $\lambda \in [0, 1]$ ) 表示. 在 MRR2 方法中, 参数拟合也是利用最小二乘方法进行拟合; 非参数方法则是对参数拟合与真实响应值的残差进行非参数拟合. 所谓参数拟合和真实响应的残差是指参数拟合无法获取的数据结构, 而本节的非参数拟合是指通过 LLR 来对上述残差进行拟合, 其混合模型为

$$\hat{\mathbf{y}}^{\text{MRR2}} = \hat{\mathbf{y}}^{\text{OLS}} + \lambda \hat{\mathbf{r}},$$

其中  $\hat{\mathbf{r}} = \mathbf{H}_r^{\text{LLR}} \mathbf{r}$ ,  $\mathbf{H}_r^{\text{LLR}}$  是指 OLS 方法得到的残差作为响应来进行非参数拟合的 LLR 平滑矩阵.  $\lambda$  是本方法的混合参数.

类似于 MRR1 中的  $\lambda$  的求法,  $\lambda \in [0, 1]$ ,  $\lambda$  的大小并不像上一节方法可以反映模型错误估计的程度, 而是指表示残差拟合需要的校正量. 当  $\lambda = 1$  时表示非参数拟合和参数拟合所占的比重一样. 渐近最优估计的  $\lambda$  表达式为  $\hat{\lambda}_{\text{opt}}^{\text{MRR2}} = \langle \hat{\mathbf{r}}, \mathbf{y} - \hat{\mathbf{y}}^{\text{OLS}} \rangle / \|\hat{\mathbf{r}}\|^2$ . 根据以往文献发现, 关于响应估计的期望, 一般认为 MRR2 的效果比 MRR1 的效果更加好. 因此, 本文将倾向于使用 MRR2 方法进行参数估计.

## 3 本文所提方法

### 3.1 考虑重复试验的非参数响应曲面建模

在多响应优化设计中, 若无法根据试验设计数据获得响应曲面模型的结构时, 采用传统的参数回归方法进行均值响应曲面建模则可能会引起较大的偏差, 即造成较大的模型误差. 鉴于上述情况, 引入非参数方法

来构建响应与试验因子之间的响应曲面模型,其具体模型为

$$y_i = h(\mathbf{x}_i) + \varepsilon_i, \quad (6)$$

其中响应预测值是由非参数拟合部分  $h(\mathbf{x}_i)$  和误差部分  $\varepsilon_i$  组成,其中  $\varepsilon_i \sim N(0, 1)$ .

若试验在每个处理(即试验方案)下都存在重复试验时,可以借鉴双曲面建模思想利用非参数方法来构建基于响应均值和方差的双响应曲面模型.结合上一节的非参数响应曲面建模知识,本文选择相对简单的局部多项式回归非参数方法,其中核函数选择常用的高斯核函数  $\kappa(u) = e^{-u^2}$ .上述核函数的带宽  $b$  可以利用式(2)通过最小化 PRESS\*\* 来获得.

针对具有重复试验数据的多响应优化问题,利用非参数模型(即式(9))来分别拟合各响应方差以及这些响应之间的协方差与试验因子  $\mathbf{x}$  的函数关系.在此基础上获得响应方差-协方差函数  $\Sigma_y(\mathbf{x})$  与试验因子  $\mathbf{x}$  之间的关系.另外,利用重复试验数据获得每个试验点的样本均值,然后利用非参数模型即式(9)拟合响应均值与试验因子之间的响应曲面模型,即均值函数  $E[\mathbf{y}(\mathbf{x})]$ .在此基础上,利用多响应的质量损失函数构建本文所提方法的优化模型,即

$$\text{Min } E[L(\mathbf{y}, \boldsymbol{\theta})] = (E[\mathbf{y}(\mathbf{x})] - \boldsymbol{\theta})^T \mathbf{C} (E[\mathbf{y}(\mathbf{x})] - \boldsymbol{\theta}) + \text{tr}[\mathbf{C} \Sigma_y(\mathbf{x})],$$

其中  $\boldsymbol{\theta}$  代表目标值矩阵,  $\mathbf{C}$  代表成本矩阵.

具体的非参数响应面方法的步骤如下:

**步骤 1** 根据 PRESS\*\* 函数代码迭代从候选列表中找到最优带宽值  $b$ .

**步骤 2** 代入期望质量损失函数,利用遗传算法参数优化,得到最佳参数水平值.

**步骤 3** 部分试验结果会受到噪声因子的影响使得结果的可靠性降低.为了对结果的可靠性评估,利用随机抽样(蒙特卡罗抽样),统计落在规格限内的样本数目,从而能够评估优化结果的可靠性.

### 3.2 结合贝叶斯抽样技术的误差修正模型

有重复试验数据的情况下,基于半参数方法的多响应曲面模型一般可假设为

$$\hat{y}_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + f(\mathbf{x}_i) + g^{\frac{1}{2}}(\mathbf{x}_i; \gamma) \varepsilon_i, \quad (7)$$

其中  $h(\mathbf{x}_i, \boldsymbol{\beta})$  代表参数方法部分,  $f(\mathbf{x}_i)$  代表的是非参数部分,  $g^{\frac{1}{2}}(\mathbf{x}_i; \gamma)$  是对方差模型的转换,  $\varepsilon_i$  是指随机误差.

然而上述做法存在三个不足之处:一是重复试验成本较高;二是试验数据的随机误差假设服从标准正态分布即  $\varepsilon_i \in N(0, 1)$  可能与实际情况并不相符,无法真实地反应模型随机误差的分布特征;三是上述模型仅仅在具有重复试验数据情况才能构造出方差模型.若试验数据没有重复的情况下,则需要进一步考虑如何构造合适的统计量来刻画响应曲面模型的随机误差部分.

借鉴参考文献[13]中利用贝叶斯抽样方法获取随机误差项的思想,本文将对式(7)中的随机误差项进行修正,其具体模型为

$$\hat{\mathbf{Y}}_{\text{new}} = \mathbf{W} \sqrt{\frac{v}{U}} + \hat{\mathbf{B}} \mathbf{z}(\mathbf{x}), \quad (8)$$

其中  $\mathbf{W} \in N(0, \mathbf{H}^{-1})$ ,  $v = N - p - q + 1$ ,  $\mathbf{H} = \frac{v \mathbf{S}^{-1}}{1 + \mathbf{z}(\mathbf{x})^T \mathbf{D}^{-1} \mathbf{z}(\mathbf{x})}$ ,  $\mathbf{S} = (\mathbf{Y} - \mathbf{Z} \hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{Z} \hat{\mathbf{B}})$ ,  $\mathbf{D} = \sum_{i=1}^n \mathbf{z}(\mathbf{x}_i) \mathbf{z}(\mathbf{x}_i)^T$ ,  $\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ , 在式(12)中,其随机误差项为  $\mathbf{W} \sqrt{v/U}$ .

在此,本文拟结合半参数建模思想和贝叶斯抽样方法,将式(8)中  $\hat{\mathbf{B}} \mathbf{z}(\mathbf{x})$  运用式(7)中半参数方法进行拟合.上述处理方法能够充分利用半参数方法的优势,如能够有效地处理小样本试验数据等.需要特别指出的是,式(11)中随机误差项  $g^{\frac{1}{2}}(\mathbf{x}_i^* T; \gamma) \varepsilon_i$  将在新模型中用式(12)的随机误差项  $\mathbf{W} \sqrt{v/U}$  所替代.利用贝叶斯抽样技术进行误差修正后的响应曲面模型为

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}^{\text{OLS}} + \lambda \hat{\mathbf{r}} + \mathbf{W} \sqrt{\frac{v}{U}}, \quad (9)$$

其中  $v$  是自由度,  $N$  是样本大小,  $\mathbf{W}$  是一个具有零均值向量和方差协方差矩阵为  $\mathbf{H}^{-1}$  的多元正态随机变量.  $\mathbf{z}(\mathbf{x}_i)$  是观测点的  $q \times 1$  的观测值, 其中  $\mathbf{X}$  是二阶因子模型矩阵,  $\mathbf{Y}$  是  $N \times p$  响应值矩阵.  $\hat{\mathbf{B}}$  是拟合系数向量,  $\mathbf{Z}$  是由  $N$  个  $\mathbf{z}(\mathbf{x}_i)$  向量形成的  $q \times N$  矩阵.  $U$  是一个自由度为  $v$  不依赖于  $\mathbf{W}$  的卡方随机变量.

在上述响应曲面模型的框架下, 结合双响应曲面建模思想<sup>[25]</sup>分别构建出响应均值和方差模型, 然后利用多元质量损失函数获得期望目标函数为

$$E[L(\hat{\mathbf{y}}, \boldsymbol{\theta})] = (E[\hat{\mathbf{y}}(\mathbf{x})] - \boldsymbol{\theta})^T \mathbf{C} (E[\hat{\mathbf{y}}(\mathbf{x})] - \boldsymbol{\theta}) + \text{tr}[\mathbf{C} \Sigma_{\hat{\mathbf{y}}}(\mathbf{x})]. \quad (10)$$

式(10)本质上是采用组合建模的思想所构建的一种新模型. 该模型整合了试验者根据经验所设定的参数模型与适应试验数据特点所估计的非参数模型二者的优势. 在此基础上, 利用贝叶斯方法对模型误差的修正则进一步提升了所构建模型的预测能力. 此外, 该模型还兼容了半参数方法和贝叶斯方法的特点, 能够有效地处理小样本试验数据的问题. 因此, 与以往的研究方法比较而言, 基于数据驱动建模思想的响应曲面模型, 能够更好地提取试验数据的特征, 构建更为精确、更加符合实际的响应曲面模型.

### 3.3 基于混合遗传算法的参数优化

由式(7)~式(10)可知, 结合半参数方法和贝叶斯抽样方法所构建的响应曲面模型往往具有高度复杂的非线性特征, 传统的优化方法(如线性规划、梯度优化等)将难以获得可靠的优化结果<sup>[28]</sup>. 为此, 借鉴文献[1, 13], 利用结合遗传算法和模式搜索法的混合遗传算法对上述目标函数即式(10)进行参数优化. 与传统的优化方法比较而言, 混合遗传算法有效地利用遗传算法良好的全局搜索能力和模式搜索方法突出的局部优化能力, 因此利用该方法能够有效地处理高度复杂的参数优化问题<sup>[29]</sup>.

### 3.4 评估优化结果的可靠性

在多响应曲面优化设计中, 由于模型不确定性以及随机误差等因素的影响, 将会导致输出响应呈现出相当大的波动. 在这种情形下, 如何衡量预测响应波动对优化结果的影响, 如何评估优化结果的可靠性成为一项亟待解决的研究课题. 为此, 一些研究者提出一些评估优化结果可靠性的研究方法. 其中, 最具有代表性的是贝叶斯后验概率方法<sup>[11]</sup>. 该方法的核心思想是在贝叶斯响应曲面建模框架下, 利用贝叶斯抽样方法计算预测响应值(即在最优参数设计值下的响应预测结果)落在其规格限内的概率, 从而据此度量优化结果的可靠性. 在给定的试验数据  $\text{data}$  和响应规格限  $A$  下, 其最优参数设计结果的贝叶斯后验概率的计算式为

$$P(\mathbf{x}) = \Pr(\hat{y}_i(\mathbf{x}) \in A | \text{data}) \approx \frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} I(\hat{y}_i^s(\mathbf{x}) \in A), \quad (11)$$

其中  $N_{\text{sim}}$  代表响应模拟抽样次数,  $I(\cdot)$  为示性函数.

当响应模拟抽样值落在规格限内时其后验概率结果  $P$  为 1, 否则为 0.

在多响应优化设计中, 在试验响应个数较多时, 式(11)将是一个非常复杂的高维数值积分问题. 此外, 在很多情况下, 其响应的密度函数往往非常复杂, 甚至难以直接获得. 因此, 在这种情况下, 通过贝叶斯模拟抽样方法将会大大地降低问题的难度和计算的工作量. 需要特别指出的是, 上述贝叶斯后验概率方法不仅适用于单响应参数优化问题, 而且也可以拓展到多响应曲面的参数优化问题. 在多响应优化设计中, 通过式(11)不仅可以获得单个响应预测值的边际后验概率值, 而且也可以获得多个响应预测值落在其对应规格限内的联合概率值.

### 3.5 本文所提方法的实施步骤

针对多响应曲面优化设计中存在的问题, 本文结合半参数方法和贝叶斯抽样模拟技术, 提出一种新的建模与优化技术, 其具体的实施步骤如下:

**步骤 1** 首先对原始数据进行普通最小二乘法(OLS)拟合分别得到关于两个响应的参数拟合值  $\hat{\mathbf{y}}^{\text{OLS}}$ , 然后通过原始响应构建残差向量, 将残差向量再利用 PRESS\*\* 函数寻求最佳的带宽值  $b$ .

**步骤 2** 根据半参数模型公式中的非参数部分, 利用 LLR 方法对上述步骤2得到的两组残差向量分别进行非参数拟合得到两组非参数拟合的拟合残差向量  $\hat{\mathbf{r}}$ .

**步骤3** 通过步骤1和步骤2中分别得到的OLS拟合值和非参数残差拟合值 $\hat{r}$ . 利用MRR2方法中的混合参数构造方法—— $\hat{\lambda}_{\text{opt}}^{\text{MRR2}}$ 函数, 分别得到对应响应的最合适的混合参数值 $\lambda$ . 结合OLS拟合值、残差拟合值 $\hat{r}$ 和混合参数 $\lambda$ 将非参数响应曲面模型扩展为半参数响应曲面模型.

**步骤4** 借鉴参考文献[13]中利用贝叶斯抽样方法获取随机误差项的思想进行误差修正, 同时利用半参数的响应建模方法对响应曲面进行重新拟合. 最后, 综合上述两个方面的改进提出一种基于误差修正的响应曲面模型, 并在此基础上利用多元质量损失函数构建出本文所提方法的优化目标函数.

**步骤5** 利用混合遗传算法对步骤4中所获得期望质量损失函数进行参数优化, 获得最优的参数设计值.

**步骤6** 利用贝叶斯后验概率方法对参数优化结果的可靠性进行评估.

## 4 实例分析

该实例来源于文献[13], 其主要目的是对一种聚合物进行参数优化设计, 以寻找最佳的设计参数值. 在该试验中, 可控因子为反应时间 $x_1$ (reaction time)、反应温度 $x_2$ (reaction temperature)和催化剂的用量 $x_3$ (amount of catalyst). 试验的输出响应是转化率 $y_1$ 和热活动 $y_2$ . 其中, 转化率为望大类型的质量特性( $y_1 \in [80, 100]$ ), 目标值 $\theta_1$ 假定为100; 而热活动为望目的质量特性( $y_2 \in [55, 66]$ ), 其目标值 $\theta_2$ 假定为57.5. 在整个试验的分析过程中, 假设回归模型中因子效应所构成的向量为 $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_3, x_1x_2, x_2x_3, x_1x_3, x_1^2, x_2^2, x_3^2)$ . 在该试验中, 试验者选择中心复合设计(central composite design, CCD)开展相关的试验, 其聚合物试验的因子计划安排、响应的输出结果如下表1所示.

表1 某聚合物的试验设计结果  
Table 1 The experimental design results of a polymer

| 试验次序 | 控制因子  |       |       | 响应    |       |
|------|-------|-------|-------|-------|-------|
|      | $x_1$ | $x_2$ | $x_3$ | $y_1$ | $y_2$ |
| 1    | -1.00 | -1.00 | -1.00 | 74    | 53.2  |
| 2    | 1.00  | -1.00 | -1.00 | 51    | 62.9  |
| 3    | -1.00 | 1.00  | -1.00 | 88    | 53.4  |
| 4    | 1.00  | 1.00  | -1.00 | 70    | 62.6  |
| 5    | -1.00 | -1.00 | 1.00  | 71    | 57.3  |
| 6    | 1.00  | -1.00 | 1.00  | 90    | 67.9  |
| 7    | -1.00 | 1.00  | 1.00  | 66    | 59.8  |
| 8    | 1.00  | 1.00  | 1.00  | 97    | 67.8  |
| 9    | -1.68 | 0.00  | 0.00  | 76    | 59.1  |
| 10   | 1.68  | 0.00  | 0.00  | 79    | 65.9  |
| 11   | 0.00  | -1.68 | 0.00  | 85    | 60.0  |
| 12   | 0.00  | 1.68  | 0.00  | 97    | 60.7  |
| 13   | 0.00  | 0.00  | -1.68 | 55    | 57.4  |
| 14   | 0.00  | 0.00  | 1.68  | 81    | 63.2  |
| 15   | 0.00  | 0.00  | 0.00  | 81    | 59.2  |
| 16   | 0.00  | 0.00  | 0.00  | 75    | 60.4  |
| 17   | 0.00  | 0.00  | 0.00  | 76    | 59.1  |
| 18   | 0.00  | 0.00  | 0.00  | 83    | 60.6  |
| 19   | 0.00  | 0.00  | 0.00  | 80    | 60.8  |
| 20   | 0.00  | 0.00  | 0.00  | 91    | 58.9  |

根据第3.5节中所提出的实施步骤, 结合上述表1所获得可控因子与输出响应的试验数据, 在此利用半参数建模方法进行响应曲面建模. 其中非参数方法中核函数选择为高斯核函数, 其最佳带宽值 $b$ (其中两个响应关于带宽的PRESS\*\*函数如图1, 图2所示)分别为0.68和0.35, 最佳混合参数值为1.

对半参数-误差修正模型  $\hat{y}^{MRR2} = \hat{y}^{OLS} + \lambda \hat{r} + \varepsilon$  进行计算得到两个响应的参数拟合部分分别为

$$\begin{aligned} \hat{y}_1^{OLS} &= 81.094 - 1.029x_1 - 4.043x_2 + 6.206x_3 + 2.125x_1x_2 + \\ &\quad 11.375x_1x_3 - 3.875x_2x_3 - 1.838x_1^2 + 2.946x_2^2 - 5.204x_3^3, \\ \hat{y}_2^{OLS} &= 59.851 + 3.586x_1 + 0.255x_2 + 2.231x_3 - 0.388x_1x_2 - \\ &\quad 0.038x_1x_3 + 0.313x_2x_3 + 0.836x_1^2 + 0.074x_2^2 + 0.056x_3^3. \end{aligned}$$

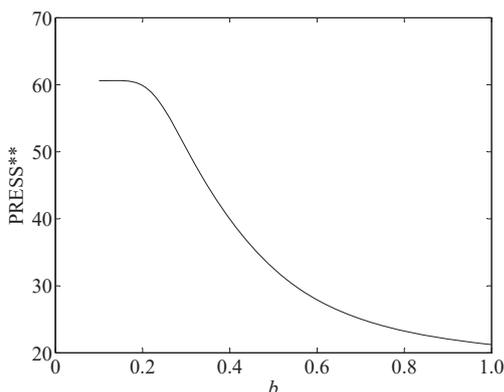


图 1 响应值  $y_1$  的 PRESS\*\* 函数图

Fig. 1 PRESS\*\* function diagram of response  $y_1$

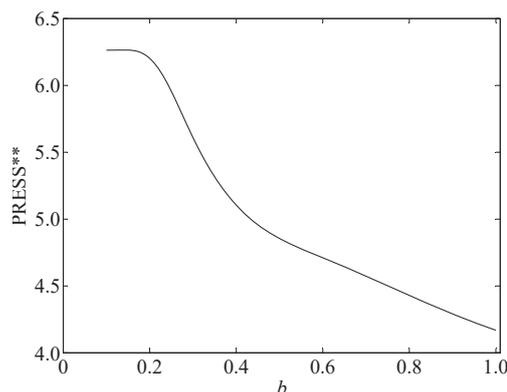


图 2 响应值  $y_2$  的 PRESS\*\* 函数图

Fig. 2 PRESS\*\* function diagram of response  $y_2$

然后, 利用第 3.3 节中所介绍的混合遗传算法对所提出的期望质量损失函数进行参数优化, 其参数优化结果为  $x_1 = -0.050, x_2 = 1.682, x_3 = -0.063$ . 同时, 在最优位置点处的期望质量损失为 13.834, 贝叶斯后验概率为 0.813 8. 若采用文献[13]方法对该聚合物试验进行分析可知, 其期望质量损失函数为 14.663 6, 同时其贝叶斯后验概率(即蒙特卡洛的抽样值落在给定规格限内的概率)为 0.602 8. 与文献[13]研究结果比较而言, 本文所提方法在期望质量损失和后验概率值(该指标反映产品符合规格的程度)均表现出一定的优势, 其主要原因是文献[13]方法事先假设固定的模型结构, 从而忽视模型结构不确定性对优化结果的影响. 正如文献[13]中曾指出的那样, 模型结构的变化可能会低估或高估其综合性能评价指标(如期望质量损失函数或贝叶斯后验概率). 此外, 为了进一步验证所提方法的有效性, 本文还将文献[11]的贝叶斯后验概率方法、文献[10]的多元质量损失函数方法、文献[20]中提到的非参数、文献[20]中提到的半参数方法运用到该聚合物试验的参数优化过程中, 其优化结果如表 2 所示. 与文献[20]中的非参数方法和半参数方法两种方法比较而言, 本文所提方法在期望质量损失函数和规格限内的概率两个评估指标上均优于上述两种方法的优化结果. 因此, 本文所提方法不仅在期望质量损失方面有所降低, 而且在考虑产品的符合性概率方面也获得了较大的提升.

此外, 为了进一步验证参数优化结果的可靠性, 在最优参数设计值下利用所构建的响应模型进行 10 000 次随机模拟抽样. 考虑到样本间的自相关性, 每间隔 4 个抽样点抽取一次, 最后获得了 2 000 次的响应抽样值, 其两个响应  $y_1$  和  $y_2$  抽样值的踪迹图如图 3 和图 4 所示.

在图 3 中, 其两条红线分别代表响应  $y_1$  的规格上限 100 和规格下限 80. 由于响应  $y_1$  为望大类型的, 因此其响应目标值假设为 100, 目标值红线和上限红线重合. 另外, 在图 4 中, 其响应  $y_2$  的规格上限和下限分别为 55 和 66. 图 4 中间的红线代表响应  $y_2$  的目标值, 即为 57.5. 从图 3 和图 4 可知, 由于随机误差等确定性因素的影响, 上述两个响应的抽样值围绕着某个确定的均值呈现上下随机波动的趋势, 且其波动幅度基本保持一致, 呈现出稳态分布的特征, 因此可以利用其响应抽样值来评估其参数优化结果的可靠性. 根据上述抽样结果, 可以计算得出两个响应抽样值落在其对应的规格限内的概率分别为 0.845 2 和 0.956 6. 上述的研究结果进一步证实参数优化结果的可靠性.

表2 优化结果指标对比表  
Table 2 Comparison table of optimization results

| 试验次序                | 响应预测值       |             | 期望质量损失(EL) | 规格限内概率值(P) |
|---------------------|-------------|-------------|------------|------------|
|                     | $\bar{y}_1$ | $\bar{y}_2$ |            |            |
| 文献[11](贝叶斯后验概率)     | 86.49       | 57.75       | 20.958 1   | 0.647 8    |
| 文献[10](多元质量损失函数)    | 93.77       | 59.98       | 11.666 2   | 0.480 6    |
| 文献[13](结合质量损失与后验概率) | 89.77       | 57.88       | 14.663 6   | 0.602 8    |
| 文献[20](非参数方法)       | 93.80       | 57.21       | 14.366 1   | 0.682 4    |
| 文献[20](半参数方法)       | 94.10       | 57.13       | 13.990 2   | 0.701 4    |
| 本文方法(半参数-误差修正方法)    | 94.43       | 57.85       | 13.834 0   | 0.813 8    |

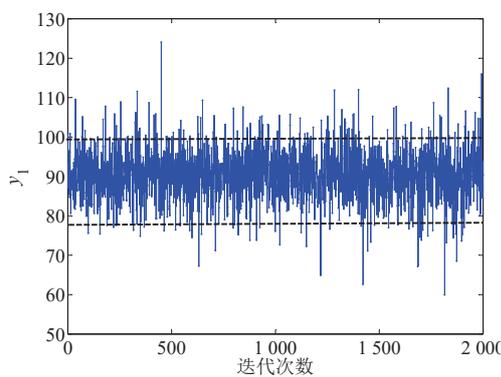


图3 响应值  $y_1$  的轨迹图

Fig. 3 Trace plot of response value  $y_1$

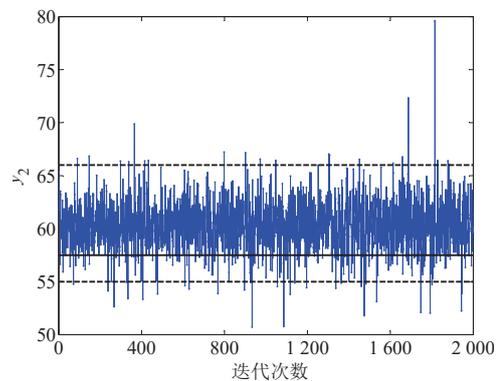


图4 响应值  $y_2$  的轨迹图

Fig. 4 Trace plot of response value  $y_2$

## 5 结束语

本文利用半参数方法建立过程响应与可控因子的响应曲面模型,并在此基础上结合贝叶斯抽样方法对所构建的响应曲面模型进行误差修正,从而获得更加精确、更加符合实际情况的响应曲面模型.此外,本文还利用贝叶斯后验概率方法进一步地评估了参数优化结果的可靠性.需要特别指出的是,若根据工程背景或专家知识能明确给出响应曲面模型结构时,则使用参数回归方法往往能够获得理想的模型估计结果.然而,若模型结构无法确定,则采用非参数或半参数方法等数据驱动方法往往能够获得更为可靠的响应曲面模型.

此外,未来的产品质量设计还需要从产品全生命周期视角全面地分析与思考.例如,在产品的设计阶段,考虑通过互联网让广大客户参与产品质量设计,让所制造出的产品全面反映客户需求,真正做到高度客户化;在产品生产阶段,考虑利用物联网与传感器、以及人工智能技术,对生产过程进行实时质量监控与反馈,及时采集生产过程中环境(如温度、湿度)、设备故障等数据信息,然后利用大数据分析技术对所收集到的上述数据进行全面地分析,探寻可能影响最终产品设计质量的关键因素并加以优化;在产品售后以及使用阶段,充分地利用互联网或物联网及时收集客户在使用中所发现的潜在质量问题,或客户在维修保养过程中存在的产品质量问题,然后结合大数据分析技术寻求潜在的原因,并在未来产品质量设计中逐步加以改进和完善<sup>[30]</sup>.因此如何结合互联网、人工智能与大数据分析技术开展产品质量设计技术与方法将是未来有待进一步研究的重要课题.

## 参考文献:

- [1] He Z, Zhu P F, Park S H. A robust desirability function method for multi-response surface optimization considering model uncertainty. *European Journal of Operational Research*, 2012, 221(1): 241-247.

- [2] Myers R H, Montgomery D C, Vining G G, et al. Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology*, 2018, 36(1): 53–77.
- [3] 汪建均, 马义中, 翟云焕. 相关多质量特性的优化设计. *管理工程学报*, 2011, 25(2): 66–73.  
Wang J J, Ma Y Z, Zhai Y H. Optimization design of correlated multiple quality characteristics. *Journal of Industrial Engineering and Engineering Management*, 2011, 25(2): 66–73. (in Chinese)
- [4] 马义中, 田 甜, 刘利平. 自相关过程协方差阵的残差 MEWMA 控制图. *系统工程学报*, 2012, 27(2): 279–286.  
Ma Y Z, Tian T, Liu L P. Residual-based MEWMA control chart for the covariance matrix of autocorrelated processes. *Journal of Systems Engineering*, 2012, 27(2): 279–286. (in Chinese)
- [5] 王 宁, 张 帅, 刘玉敏, 等. 多工序串并联制造过程关键质量特性识别. *系统工程学报*, 2019, 34(6): 855–866.  
Wang L, Zhang S, Liu Y M, et al. Identification the key quality characteristics in multistage series-parallel manufacturing process. *Journal of Systems Engineering*, 2019, 34(6): 855–866. (in Chinese)
- [6] Wang J J, Ma Y Z, Ouyang L H, et al. Bayesian modeling and optimization for multi-response surfaces. *Computers & Industrial Engineering*, 2020, 142, 106357. <https://doi.org/10.1016/j.cie.2020.106357>.
- [7] Wang J J, Ma Y Z, Ouyang L H, et al. A new Bayesian approach to multi-response surface optimization integrating loss function with posterior probability. *European Journal of Operational Research*, 2016, 249(1): 231–237.
- [8] He Z, Wang J, Oh J, et al. Robust optimization for multiple responses using response surface methodology. *Applied Stochastic Models in Business and Industry*, 2010, 26(2): 157–171.
- [9] Ko Y H, Kim K J, Jun C H. A new loss function-based method for multiresponse optimization. *Journal of Quality Technology*, 2018, 37(1): 50–59.
- [10] Box G E P, Draper N R. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987.
- [11] Peterson J J. A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology*, 2018, 36(2): 139–153.
- [12] Peterson J J, Miró-Quesada G, del Castillo E. A bayesian reliability approach to multiple response optimization with seemingly unrelated regression models. *Quality Technology & Quantitative Management*, 2016, 6(4): 353–369.
- [13] 汪建均, 马义中, 欧阳林寒, 等. 多响应稳健参数设计的贝叶斯建模与优化. *管理科学学报*, 2016, 19(2): 85–94.  
Wang J J, Ma Y Z, Ouyang L H, et al. Bayesian modeling and optimization of multi-response robust parameter design. *Journal of Management Sciences in China*, 2016, 19(2): 85–94. (in Chinese)
- [14] Min J Y. *Off-line quality control by robust parameter design*. Manhattan: Iowa State University, 2005.
- [15] Vining G G, Bohn L L. Response surfaces for the mean and variance using a nonparametric approach. *Journal of Quality Technology*, 1998, 30(3): 282–291.
- [16] Ouyang L H, Wan L Q, Park C, et al. Ensemble RBF modeling technique for quality design. *Journal of Management Science and Engineering*, 2019, 4(2): 105–118.
- [17] Fox J, Weisberg S. *An R Companion to Applied Regression*. Thousand Oaks: Sage, 2019.
- [18] Lim Y, Oh H S. Confidence intervals for nonparametric quantile regression: An emphasis on smoothing splines approach. *Australian & New Zealand Journal of Statistics*, 2017, 59(4): 527–543.
- [19] Wan W. *Semi-parametric techniques for multi-response optimization*. Blacksburg: Virginia Polytechnic Institute & State University, 2007.
- [20] Pickle S M, Robinson T J, Birch J B, et al. A semi-parametric approach to robust parameter design. *Journal of Statistical Planning and Inference*, 2008, 138(1): 114–131.
- [21] Mays J E, Birch J B, Alden Starnes B. Model robust regression: Combining parametric nonparametric and semi-parametric methods. *Journal of Nonparametric Statistics*, 2001, 13(2): 245–277.
- [22] Mays J E, Birch J B, Einsporn R L. An overview of model-robust regression. *Journal of Statistical Computation and Simulation*, 2000, 66(1): 79–100.
- [23] Mays J E, Birch J B. Smoothing for small samples with model misspecification: Nonparametric and semi-parametric concerns. *Journal of Applied Statistics*, 2002, 29(7): 1023–1045.
- [24] Wan W, Birch J B. A semi-parametric technique for the multi-response optimization problem. *Quality and Reliability Engineering International*, 2011, 27(1): 47–59.
- [25] Anderson-Cook C M, Prewitt K. Some guidelines for using nonparametric methods for modeling data from response surface designs. *Journal of Modern Applied Statistical Methods*, 2005, 4(1): 106–119.

- [26] Simonoff J S. Smoothing Methods in Statistics. New York: Springer Series in Statistics, 1996.
- [27] Robinson T J, Birch J B, Starnes B A. A semi-parametric approach to dual modeling when no replication exists. *Journal of Statistical Planning and Inference*, 2010, 140(10): 2860–2869.
- [28] 薛丽, 何桢. 二项分布下 VSI EWMA 控制图的经济设计. *系统工程学报*, 2020, 35(5): 711–720.  
Xue L, He Z. Economic design of variable sampling intervals EWMA control chart under binomial distributions. *Journal of Systems Engineering*, 2020, 35(5): 711–720. (in Chinese)
- [29] 汪建均, 马义中. 基于 GLM 的双响应曲面法及其稳健设计. *系统工程与电子技术*, 2012, 34(11): 2306–2311.  
Wang J J, Ma Y Z. Dual response surface methodology based on generalized linear models and its application on robust design. *Systems Engineering and Electronics*, 2012, 34(11): 2306–2311. (in Chinese)
- [30] 刘海燕, 王宗水, 汪寿阳. 我国系统科学与工程研究的演化与发展. *系统工程学报*, 2017, 32(3): 289–304.  
Liu H Y, Wang Z S, Wang S Y. Evolution and development of system science and engineering research. *Journal of Systems Engineering*, 2017, 32(3): 289–304. (in Chinese)

### 作者简介:

汪建均(1977—), 男, 湖南慈利人, 博士, 教授, 博士生导师, 研究方向: 质量管理与质量工程, 工业工程, 应用统计学, Email: jjwang@njust.edu.cn;

郜婷玉(1995—), 女, 安徽铜陵人, 硕士生, 研究方向: 质量管理与质量工程, Email: 1721922493@qq.com;

杨世娟(1988—), 女, 安徽安庆人, 博士生, 研究方法: 质量管理与质量工程, Email: yang3025594252@163.com.

\*\*\*\*\*

(上接第 589 页)

- [20] Cui Y, Wang X. Detecting one-mode communities in bipartite networks by bipartite clustering triangular. *Physica A: Statistical Mechanics and its Applications*, 2016, 457: 307–315.
- [21] Khazaei E, Alimohammadi A. An automatic user grouping model for a group recommender system in location-based social networks. *ISPRS International Journal of Geo-Information*, 2018, 7(2): 67–84.
- [22] Zhou X, Wu B, Jin Q. Analysis of user network and correlation for community discovery based on topic-aware similarity and behavioral influence. *IEEE Transactions on Human-Machine Systems*, 2017, 48(6): 559–571.
- [23] Zhu J, Ma S, Zhang H, et al. Incorporating user grouping into retweeting behavior modeling // *International Conference on Database Systems for Advanced Applications*. Berlin: Springer, 2018: 474–490.
- [24] Schantl J, Kaiser R, Wagner C, et al. The utility of social and topical factors in anticipating repliers in twitter conversations // *Proceedings of the 5th Annual ACM Web Science Conference*. New York: ACM, 2013: 376–385.
- [25] Yuan N J, Zhong Y, Zhang F, et al. Who will reply to/retweet this tweet. The dynamics of intimacy from online social interactions // *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. New York: ACM, 2016: 3–12.
- [26] Saracco F, Straka M J, Di Clemente R, et al. Inferring monopartite projections of bipartite networks: An entropy-based approach. *New Journal of Physics*, 2017, 19(5): 053022.
- [27] Becatti C, Caldarelli G, Lambiotte R, et al. Extracting significant signal of news consumption from social networks: The case of Twitter in Italian political elections. *Palgrave Communications*, 2019, 5(1): 91–106.
- [28] 刘彩虹, 郭崇慧. 移动情境感知环境下的用户行为模式挖掘算法研究. *系统工程学报*, 2020, 35(4): 433–445.  
Liu C H, Guo C H. Algorithm for user behavior pattern mining in mobile context-aware environments. *Journal of Systems Engineering*, 2020, 35(4): 433–445. (in Chinese)

### 作者简介:

张亚茹(1994—), 女, 河北张家口人, 博士生, 研究方向: 系统集成, 大数据与网络舆情, 自然语言处理, 社会网络分析, Email: zhangyaru@amss.ac.cn;

唐锡晋(1967—), 女, 江苏无锡人, 博士, 研究员, 博士生导师, 研究方向: 系统集成, 知识管理, 决策支持系统, 互联网舆情, Email: xjtang@iss.ac.cn.