

doi: 10.19920/j.cnki.jmsc.2026.06.006

# 面向质量设计的计算机试验元建模研究<sup>①</sup>

——基于统计检验视角

欧阳林寒<sup>1,2</sup>, 陶宝平<sup>1</sup>, 何 桢<sup>3\*</sup>

(1. 南京航空航天大学经济与管理学院, 南京 211106; 2. 南京航空航天大学“制造系统数智管理与低碳运营”江苏省高校哲学社会科学实验室, 南京 211106; 3. 天津大学管理与经济学部, 天津 300072)

摘要: 针对质量设计中的计算机试验元建模过程, 本研究综合考虑质量变量不确定性与模型结构不确定性, 提出了基于统计检验的贝叶斯 Kriging 模型, 为持续质量改进奠定基础。该方法以贝叶斯层次模型与因子效应原则为基础, 不仅筛选出了 Kriging 全局趋势模型中的显著性变量, 而且从统计检验角度考虑了模型的有效性。首先, 在 Kriging 模型中引入二元变量指示器, 通过在参数先验分布中融合因子效应原则, 厘清变量间关系同时降低变量空间维度; 其次, 运用马尔科夫链蒙特卡洛模拟估计模型后验概率, 得到显著变量下的候选模型集合; 接着, 结合多重假设检验与 Bootstrap 预测方差修正, 对候选模型有效性进行分析, 综合考虑模型的有效性与泛化能力得到最佳模型; 最后, 通过仿真试验验证了所提方法在不同显著性水平和样本规模下均表现较好的预测效果, 同时两类工业情景下的建模结果表明, 所提方法在差异性后验概率与无差异后验概率下均能有效地识别出显著性变量。

关键词: 质量设计; Kriging 模型; 贝叶斯层次模型; 变量筛选; 统计检验

中图分类号: F253.3; O213.1 文献标识码: A 文章编号: 1007-9807(2026)06-0075-16

## 0 引言

2023 年 2 月, 中共中央、国务院印发的《质量强国建设纲要》中明确指出, “加快质量技术创新应用, 推进质量设计、试验检测、可靠性工程等先进质量技术的研发应用”。无论是在摆位, 还是质量技术本身, “质量设计”得到了极大地关注。通常, 设计参数与产品质量特性之间存在着复杂非线性的关系, 如何根据输入的变量与输出的响应构建精确有效的质量模型, 是开展质量设计工作的关键。随着质量 4.0 时代的到来, 新一代信息技术深度融入到产品制造管理过程, 促使质量管理机理与方法发生了改变<sup>[1,2]</sup>。工信部在 2021 年末编制的《制造业质量管理数字化实施指南(试

行)》中指出, 企业应充分运用数字分析建模与仿真工具提高质量水平。因此在质量 4.0 时代, 数据模型驱动下的质量预测与设计已成为质量管理转型的重要方向<sup>[3]</sup>。

在新时代发展背景下, 基于计算机试验设计的元建模技术给质量管理带来了新的思路。试验设计下的质量提升策略最早在 1926 年由 Fisher 提出, 该方法被视为 20 世纪上半叶最重要的科学发展之一<sup>[4]</sup>。历经近一个世纪的发展, 美国工程院院士、著名质量专家 Jeff Wu 指出了“后 Fisher 时代”物理试验设计向计算机试验设计转变的必然性和优势性<sup>[5]</sup>。作为计算机试验设计代表模型之一的 Kriging 模型, 由于其灵活性强、精度高等优点, 已成为当前质量设计领域中的关键工具之

① 收稿日期: 2023-02-09; 修订日期: 2024-08-06。

基金项目: 国家自然科学基金资助项目(72471112; 72072089; 72231005; 72261147706; 92467302)。

通讯作者: 何 桢(1967—), 男, 河南濮阳人, 博士, 教授, 博士生导师。Email: zhhe@tju.edu.cn

—<sup>[6,7]</sup>. 在质量建模过程中,工艺参数随现代制造过程复杂化而日趋增多<sup>[8-10]</sup>,如在增材制造熔体静电纺丝过程中,影响纺丝成型质量的变量包括熔体温度、熔体流量以及高压静电场等诸多因素<sup>[11]</sup>;在航空离心压气机质量改进中,需要考虑叶轮外径、间隙以及压力面厚度等众多变量影响<sup>[12]</sup>. 此外,当输入与输出间存在复杂非线性关系时,除需考虑变量主效应的影响,往往还需将交互效应和二次效应纳入建模考虑中<sup>[13,14]</sup>. 但是随着影响因子数量的增加,效应的数量会以指数型增加,其中大量效应是冗余. 冗余的效应不仅会增加建模工作量,也会增大模型过拟合风险. 而依据统计学中的效应稀疏原理<sup>[15]</sup>与管理学中的二八定律,众多效应中真正影响输出的因素往往只占少部分,并且在能够较好地解释数据的基础上,工程人员倾向于使用“少数”效应的模型对制造/服务过程进行分析<sup>[16,17]</sup>.

此外,由于模型只是实际问题的简化近似,为确保模型在后续质量改进活动中的可靠性,需将模型结构不确定性纳入考虑中,避免因无效模型导致错误决策. 现有国内外学者在研究变量筛选时大多数以变量不确定性为研究对象,忽略了模型结构不确定性影响,同时在此基础上从统计检验角度考虑模型有效性缺乏相关理论研究. 因此,基于计算机试验设计质量模型的研究现状,融合因子效应原则与统计检验理论,考虑质量变量和模型结构双重筛选框架下的质量设计具有重要的现实意义. 为此,本研究提出在贝叶斯 Kriging 建模框架下同时考虑变量不确定性与模型结构有效性,首先,引入二元变量指示器,并将因子效应原则融入参数先验分布中;然后,运用马尔科夫链蒙特卡洛模拟估计模型后验概率,识别出显著模型;最后,运用基于 Bootstrap 方差修正的多重假设检验对候选模型有效性进行分析,综合模型有效性与泛化能力选择最佳质量模型.

## 1 国内外研究现状

近年来,产品生产制造过程日趋复杂,随之带

来了高昂的物理试验成本,以计算机元建模为主的质量设计问题逐渐引起了广泛关注,其中针对 Kriging 元建模过程中关键因子识别的精确度与效率问题,学术界展开了深入研究. Joseph 等<sup>[18]</sup>首次将贝叶斯前向筛选法融入 Kriging 模型构建中,提出基于变量选择的盲 Kriging 模型,综合分析了显著变量与非显著变量对质量预测的影响. Huang 等<sup>[19]</sup>在贝叶斯框架下,从变量筛选的角度考虑了分层 Kriging 模型的构建,并从理论上证明了筛选算法的收敛性. Deng 等<sup>[20]</sup>通过对均值函数系数赋予贝叶斯稀疏先验,实现模型关键变量识别与变量系数自适应估计. 与贝叶斯方法不同,惩罚似然函数通过加入惩罚项,权衡模型的拟合精度与变量稀疏性. Hung<sup>[21]</sup>首次从惩罚似然函数的角度考虑了 Kriging 建模过程中均值函数的变量降维问题. Zhang 等<sup>[22]</sup>采用正则化的方法对均值函数和相关函数进行双重惩罚,并讨论了质量变量对两类函数的影响. 随后, Park<sup>[23]</sup>通过 Lasso 惩罚函数更新极大似然估计方法,基于交叉验证的最小角回归算法分析数据维度对 Kriging 模型变量筛选效率和稳健性能的影响. 此外,相关学者提出从高维矩阵处理的角度提升变量筛选的效率, Shao 等<sup>[24]</sup>提出了稀疏 Kriging 质量模型,该工作的创新之处在于其应用 Cholesky 分解降低了相关矩阵的维度,通过循环迭代的方式确定均值函数中的显著变量. 林成龙等<sup>[25]</sup>针对复杂系统工程中工艺参数过多导致 Kriging 建模效率低下的问题,提出利用偏最小二乘法将高维矩阵投影到低维空间. 同时, Park 等<sup>[26]</sup>通过梯度下降算法优化稀疏投影矩阵的分解工作,提高了质量变量的筛选效率.

此外,元模型只是对实际问题的近似,在应用于质量设计之前须确保其准确有效,然而由于各方法在统计指标和筛选机制等方面的差异,导致不同方法下的模型结构对质量预测存在高估或低估现象. Kleijnen 和 Beers<sup>[27]</sup>首次从统计检验的角度分析了模型结构差异对质量预测的影响,并构建了基于交叉验证的 Kriging 模型筛选法. Demay 等<sup>[28]</sup>指出模型结构对质量预测方差有显著影响,

并提出了基于估计方差准则的模型结构确定方法。基于组合策略的 Kriging 模型预测研究明晰了模型结构不确定性的变异源。Ba 和 Joseph<sup>[29]</sup> 从不同 Kriging 模型结构的角度, 讨论了组合结构在试验设计稀疏情形下的优势, 并从理论上进一步证明了组合策略下参数估计的一致性。欧阳林寒等<sup>[8]</sup> 讨论了核函数结构差异下的 Kriging 模型预测精度问题, 采用 Bootstrap 方法分析了各核函数的性能, 并通过组合核函数降低模型结构对质量预测的影响。肖甜丽等<sup>[30]</sup> 针对核函数选择的不确定性, 提出利用随机搜索变量选择法选择组合的核函数类型, 并通过 K 均值聚类法确定了组合的 Kriging 模型结构。万良琪和欧阳林寒<sup>[31]</sup> 进一步研究了具有多个核函数的 Kriging 组合建模问题, 提出了基于 0-1 规划的筛选策略, 确定了基于组合核函数的 Kriging 模型结构。考虑到均值函数变异源对 Kriging 建模的影响, Zhang 和 Taflanidis<sup>[32]</sup> 从贝叶斯模型平均的角度, 以各个模型的后验概率为权重, 综合不同显著变量集合下 Kriging 模型的预测效果, 提高质量预测结果的鲁棒性。

综上, 均值函数或核函数的构成形式和预测的不同组合策略均为影响 Kriging 模型预测性能的重要因素, 国内外相关研究已取得较为丰富的成果, 但仍存在如下不足和局限: 第一, 研究均值函数或核函数构成形式的文献均假设各变量间是独立同分布, 未将因子效应原则融入变量筛选过程; 第二, 组合策略中各模型结构均被假设为有效的, 缺乏对模型的有效检验, 难以保证组合策略能够显著提升模型的预测性能; 第三, 相关文献虽从统计检验的角度开展了 Kriging 模型的研究, 但未将变量筛选和组合策略纳入统计检验框架下进行系统研究。

鉴于上述不足, 本研究尝试进行有益的补充, 基于统计检验的视角, 结合因子效应原则构建分层先验, 建立融变量筛选与模型选择的一体化 Kriging 建模方法。主要贡献为: 1) 在贝叶斯层次模型框架下, 构建了面向 Kriging 全局趋势模型的变量筛选算法。该算法的优势在于因子效应原则不仅可以厘清变量间的逻辑关系, 而且能够显著

提高变量筛选效率; 2) 针对基于贝叶斯后验概率确定的 Kriging 模型结构集, 提出了尺度独立的统计检验算法, 可以避免仅依靠后验概率指标带来的模型选择“错误”, 即从统计视角看, 后验概率高的模型并不一定是有效模型; 3) 探索了单一 Kriging 模型结构和组合 Kriging 模型结构孰优孰劣的比较条件, 并具体分析各模型结构在后验概率指标上的差异性对预测性能的影响。

## 2 面向质量设计的 Kriging 建模

Kriging 模型将工艺参数与质量特性之间的复杂非线性关系视为黑箱函数, 该黑箱函数将生产过程看作是某个高斯随机过程的具体实现。Kriging 模型主要包括全局趋势模型(均值函数)与平稳高斯过程模型两个部分

$$y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta} + Z(\boldsymbol{x}) \quad (1)$$

其中全局趋势模型  $\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}$  用于近似随机过程均值; 随机项  $Z(\boldsymbol{x})$  服从均值为 0, 方差为  $\sigma^2$  的平稳高斯过程, 用于模型空间插值。给定设计空间任意两点  $\boldsymbol{x}_i, \boldsymbol{x}_j$ , 响应  $y(\boldsymbol{x}_i)$  与  $y(\boldsymbol{x}_j)$  的协方差为

$$\text{Cov}(y(\boldsymbol{x}_i), y(\boldsymbol{x}_j)) = \sigma^2 R(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (2)$$

其中  $\sigma^2$  为过程方差,  $R(\boldsymbol{x}_i, \boldsymbol{x}_j)$  为相关函数, 采取常用的高斯核函数度量样本间相关关系, 则在设计空间  $\mathbb{R}^d$  中  $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})^T$ ,  $\boldsymbol{x}_i$  与  $\boldsymbol{x}_j$  两点的相关函数如下

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\sum_{m=1}^d \theta_m (x_{im} - x_{jm})^2\right) \quad (3)$$

采取最大似然估计对模型中超参数进行优化, 取似然函数  $L(\boldsymbol{\theta})$  的对数如式(4)所示

$$\ln L(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\boldsymbol{R}|) - \frac{(\boldsymbol{y} - \hat{\boldsymbol{F}}\boldsymbol{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{F}}\boldsymbol{\beta})}{2\hat{\sigma}^2} \quad (4)$$

其中  $\boldsymbol{F} = [f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n)]^T$ ,  $f(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})]^T$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$ ,  $\boldsymbol{\beta}$  与  $\sigma^2$  的估计均可以利用  $\boldsymbol{\theta}$  表示, 因此似然函数仅为  $\boldsymbol{\theta}$  的函数, 通过优化算法全局寻优得到  $\boldsymbol{\theta}$  最优解。

类似于线性回归模型, Kriging 模型的全局趋势模型  $\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}$  中通常包括常数项、变量一次效

应、二次效应和交互效应等. 其中,常见的普通 Kriging 模型(Ordinary Kriging, OK) 采取常数对全局趋势模型进行近似: 结构较为简单易于使用,但在具有变化趋势的工业数据中表现较差,同时在参数  $\theta$  发生变化时稳健性较差<sup>[32]</sup>. 相比之下,泛 Kriging 模型(Universal Kriging, UK) 虽然考虑了各类变量效应对响应变化的影响,但模型中冗余变量增多,且在样本数量少的情况下,采取传统工具包无法进行参数估计<sup>[33]</sup>. 鉴于此,本研究在 UK 模型框架下,基于变量筛选思想和层级贝叶斯方法提出了改进的贝叶斯 Kriging 模型(Bayesian Kriging, BK)

$$y(x) = v(x)^T \beta + Z(x) \tag{5}$$

其中  $v(x)$  中仅考虑了对输出响应影响显著的  $s$  个变量. 对于待预测点  $x_0$ , 在均方误差最小的情况下, BK 模型给出该点处  $y(x_0)$  的最优线性无偏估计值

$$\hat{y}(x_0) = v(x_0)^T \hat{\beta} + r(x_0)^T R^{-1}(y - V\hat{\beta}) \tag{6}$$

与预测方差

$$\hat{s}(x_0) = \hat{\sigma}^2(1 - r(x_0)^T R^{-1}r(x_0) + \xi^T(V^T R^{-1}V)^{-1}\xi) \tag{7}$$

其中  $\xi = V^T R^{-1}r(x_0) - v(x_0)$ ,  $R$  为样本间相关矩阵,  $r(x_0)$  为  $x_0$  与训练样本相关向量,  $V = [v(x_1) \dots v(x_n)]^T$ ,  $v(x) = [v_1(x) \dots v_s(x)]^T$ ,  $v_i(x)$   $i = 1, \dots, s$  选取将在第 3 节讨论.

### 3 层级贝叶斯 Kriging 模型构建与统计检验

#### 3.1 层级贝叶斯 Kriging 模型构建

针对 Kriging 全局趋势模型显著变量  $v(x)$  选择问题, 本节在层级贝叶斯框架下, 融合因子效应原则, 构建层级贝叶斯 Kriging 模型实现显著变量, 在有效把握数据变化规律的同时降低模型复杂度.

值得指出的是, Kriging 模型中的全局趋势模型与多元线性回归模型中的线性回归项两者虽然形式相同, 但不同于后者中白噪声误差假设, Kriging 模型考虑了样本间相关关系, 因此两者在

参数统计推断与变量筛选中存在三点差异: 1) 在响应先验设定方面, Kriging 模型假设不同样本存在相关性, 响应先验分布中协方差阵为非对角矩阵<sup>[34, 35]</sup>, 而多元线性回归模型假设随机项  $\varepsilon$  为白噪声, 输出响应服从协方差为对角矩阵的多元正态分布; 2) 在模型参数设定方面, Kriging 模型引入相关函数度量样本间相关关系, 除与多元线性回归模型中共有的回归系数  $\beta$ 、方差  $\sigma^2$  外, 还包括相关函数中的尺度参数  $\theta$ ; 3) 在参数后验推断方面, 不同于多元线性回归中未考虑样本间相关关系对建模的影响, Kriging 模型在参数  $\beta$ 、参数  $\gamma$  以及参数  $\sigma^2$  的后验分布推导过程中涉及样本协方差矩阵  $\sigma^2 R$  计算, 因此两者后验分布存在差异.

#### 3.1.1 模型先验信息选择

在贝叶斯分析中先验分布的选取影响模型能否有效识别出显著变量, 因此合理选择参数先验分布尤为重要. Kriging 模型在处理残差项  $Z(x)$  时考虑样本间相关性, 将残差项视为服从平稳高斯过程  $Z(x) \sim GP(0, \sigma^2 R)$ , 基于平稳高斯过程假设, 对模型先验取多元正态分布

$$[y \mid \beta, \sigma^2, \gamma, \theta] \sim GP(F\beta, \sigma^2 R) \tag{8}$$

为直观识别出显著变量, 引入二元指示变量  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)^T$ , 其中

$$\gamma_i = \begin{cases} 1 & f_i(x) \text{ 为显著变量} \\ 0 & f_i(x) \text{ 为非显著变量} \end{cases} \quad i = 1, 2, \dots, k \tag{9}$$

在引入指示变量  $\gamma$  的基础上, 对每个变量系数  $\beta_i$  赋予均值为 0 的混合先验分布

$$[\beta_i \mid \gamma_i, \sigma^2] \sim \gamma_i N(0, c_i^2 \tau_i^2 \sigma^2) + (1 - \gamma_i) N(0, \tau_i^2 \sigma^2) \tag{10}$$

在该分布中使参数  $\tau_i$  取值为极小的数, 参数  $c_i$  取值为极大的数, 架构指示变量  $\gamma$  与参数  $\beta$  之间的联系. 图 1 直观地反映了  $\gamma_i$  取值与  $\beta_i$  的关系: 当  $\gamma_i = 0$  即变量  $f_i(x)$  不显著时,  $[\beta_i \mid \gamma_i, \sigma^2] \sim N(0, \tau_i^2 \sigma^2)$ , 此时  $\beta_i$  取值集中在 0 的附近; 相反, 当  $\gamma_i = 1$  即变量显著时,  $[\beta_i \mid \gamma_i, \sigma^2] \sim N(0, c_i^2 \tau_i^2 \sigma^2)$ ,  $\beta_i$  取非零值的可能性较大. 参考 Chipman 等<sup>[36]</sup>的参数选取经验, 取  $\tau_i = 1/(3 \times \Delta f_i(x))$ ,  $c_i = 10$ .

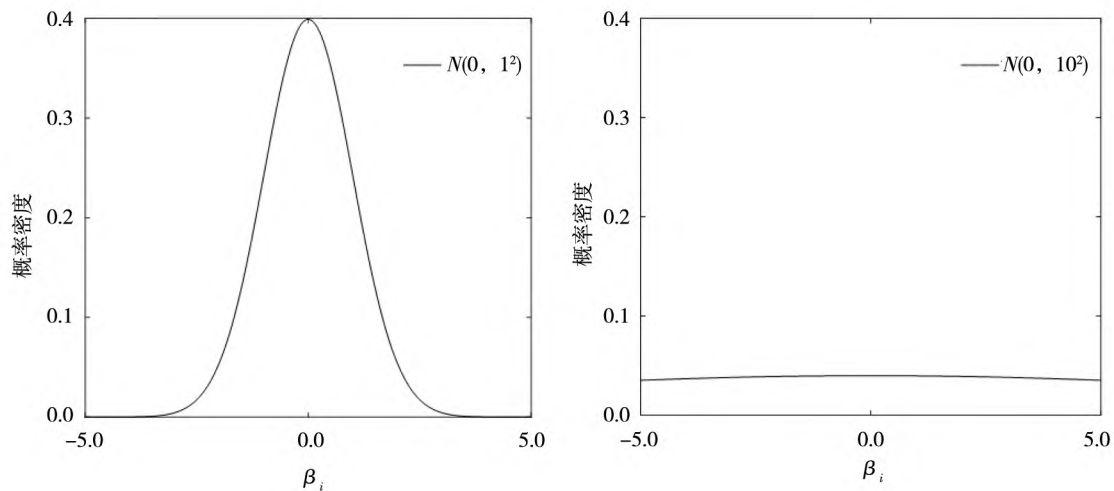


图 1 混合正态分布

Fig.1 The mixture normal distributions

二元指示变量  $\gamma_i$  先验可以视为伯努利分布，根据 Wu 和 Hamada 的试验设计原则<sup>[37]</sup>，模型变量筛选中需要考虑条件独立性原则与效应遗传原则，否则预测模型可能偏离真实模型，然而以往大多研究都忽略了这一点。条件独立性原则指出给定主效应，交互效应之间相互独立；而效应遗传原则强调当交互效应显著时，其亲本因子中必有一个显著。为直观地体现条件独立性原则与效应遗传原则下主效应与交互效应间相关关系，对  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)^T$  进行细分，以  $\gamma_{x_i}, 1 \leq i \leq d$  表示一次效应指示变量，以  $\gamma_{x_i}, d+1 \leq i \leq 2d$  表示二次效应指示变量，以  $\gamma_{x_i x_j}, 1 \leq i \neq j \leq 2d$  表示交互效应指示变量。则依据条件独立性原则，给定主效应  $\gamma_{x_i}, \gamma_{x_j}, \gamma_{x_l}$ ，交互项  $\gamma_{x_i x_j}, \gamma_{x_i x_l}, \gamma_{x_j x_l}$  之间相互独立；在效应遗传原则约束下，指示变量  $\gamma$  的先验满足

$$\pi(\gamma_{x_i x_j} = 1 \mid \gamma_{x_i}, \gamma_{x_j}) = \begin{cases} p_{11}, (\gamma_{x_i}, \gamma_{x_j}) = (1, 1) \\ p_{01}, (\gamma_{x_i}, \gamma_{x_j}) = (0, 1) \\ p_{10}, (\gamma_{x_i}, \gamma_{x_j}) = (1, 0) \\ p_{00}, (\gamma_{x_i}, \gamma_{x_j}) = (0, 0) \end{cases} \quad (11)$$

其中  $p_{00} < p_{01} = p_{10} < p_{11}$ 。综合条件独立性原则与效应遗传原则，变量  $\gamma$  先验分布满足

$$\pi(\gamma) = \prod_{i=1}^{2d} \pi(\gamma_{x_i}) \prod_{i,j=1, i \neq j}^{2d} \pi(\gamma_{x_i x_j} \mid \gamma_{x_i}, \gamma_{x_j}) \quad (12)$$

对于模型中的过程方差  $\sigma^2$ ，采取无信息先验  $\pi(\sigma^2) \propto 1/\sigma^2$ 。针对非参数部分  $Z(x)$ ，给定核参数

$\theta_i$  取值上下限，采取均匀分布  $\pi(\theta_i) \sim U(a_i, b_i)$  作为  $\theta_i$  的先验分布， $i = 1, \dots, d$ 。其中超参数  $a_i, b_i$  为  $\theta_i$  取值的下界与上界。

### 3.1.2 模型后验概率分析

基于 3.1.1 小节中讨论的各参数先验信息，在 Kriging 模型假设下展开贝叶斯分析，对各个参数的后验分布进行推断，通过计算参数的后验概率识别出候选变量集中的显著变量。命题 1 列出了各个参数的后验分布。

命题 1 根据参数  $[\beta, \sigma^2, \gamma, \theta]$  的先验信息，通过贝叶斯分析得到模型中各个参数的后验分布如下

$$[\beta \mid \sigma^2, \gamma, \theta, y] \sim N(\tilde{A}^{-1} \tilde{B} \tilde{A}^{-1} \sigma^2) \quad (13)$$

$$[\sigma^2 \mid \gamma, \theta, y] \sim IG(n/2, (y^T R^{-1} y - \tilde{B}^T \tilde{A}^{-1} \tilde{B})/2) \quad (14)$$

$$p(\theta \mid \gamma, y) \propto (y^T R^{-1} y - \tilde{B}^T \tilde{A}^{-1} \tilde{B})^{-\frac{n}{2}} \mid R \mid^{-\frac{1}{2}} \mid \tilde{A} \mid^{-\frac{1}{2}} \pi(\theta) \quad (15)$$

$$p(\gamma \mid \theta, y) \propto (y^T R^{-1} y - \tilde{B}^T \tilde{A}^{-1} \tilde{B})^{-\frac{n}{2}} \mid D \mid^{-\frac{1}{2}} \mid \tilde{A} \mid^{-\frac{1}{2}} \pi(\gamma) \quad (16)$$

其中  $D = \text{diag}((c_1^{\gamma_1} \tau_1)^2, \dots, (c_k^{\gamma_k} \tau_k)^2)$ ， $\tilde{A} = F^T \times R^{-1} F + D^{-1}$ ， $\tilde{B} = F^T R^{-1} y$ 。针对向量  $\gamma$  的后验分布难以处理与计算这一问题，根据条件概率原则依次分析其中每个分量  $\gamma_i = 1$  概率

$$\gamma_i \sim B\left(1, \frac{p(\gamma_i = 1 | \gamma_{-i} | \theta)}{p(\gamma_i = 1 | \gamma_{-i} | \theta) + p(\gamma_i = 0 | \gamma_{-i} | \theta)}\right) \tag{17}$$

其中  $\gamma_{-i}$  为去除  $\gamma_i$  后的指示向量

$$\gamma_{-i} = \begin{cases} (\gamma_2, \dots, \gamma_k)^T, & i = 1 \\ (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_k)^T, & i \neq 1, k \\ (\gamma_1, \dots, \gamma_{k-1})^T, & i = k \end{cases} \tag{18}$$

结合因子效应原则与条件独立性原则,基于指示向量  $\gamma$  后验分布可得到每个二元指示变量  $\gamma_i = 1$  的概率.

考虑到变量数量较多时,直接计算候选模型后验概率较为耗时,采取马尔科夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)模拟方法从参数后验分布中逐代采样生成马尔科夫链  $\{\sigma_{(0)}^2, \beta_{(0)}, \gamma_{(0)}, \theta_{(0)}, \dots, \sigma_{(T)}^2, \beta_{(T)}, \gamma_{(T)}, \theta_{(T)}\}$ , 动态模拟模型参数  $[\beta, \sigma^2, \gamma, \theta]$  的后验概率. 由于指示变量  $\gamma$  中包含着变量筛选的信息,因此通过研究  $\gamma$  的样本序列  $\{\gamma_{(0)}, \gamma_{(1)}, \dots, \gamma_{(T)}\}$ , 利用  $\gamma$  经过  $T$  次迭代在样本序列中出现的频率可以估计模型的后验概率. 在一组显著变量集合  $f_s(x) = [f_{s_1}(x), \dots, f_{s_m}(x)]$ ,  $1 \leq s_m \leq k$  下, Kriging 模型  $M_{\gamma_m}$  后验概率为

$$P(M_{\gamma_m}) = \frac{1}{(T-B)/s} \sum_{t=0}^{(T-B)/s-1} I(\gamma_{(ts+B+1)} = \gamma^m) \tag{19}$$

其中  $B$  为燃烧期长度,  $s$  为取样间隔以弱化参数值间自相关性,最终得到含有  $\lfloor (T-B)/s \rfloor$  个样本的序列,用于估计模型的后验概率. 算法 1 列出了应用 MCMC 识别 Kriging 显著变量的具体过程,其中  $\gamma_{i(t)}$  为第  $t$  次迭代过程中指示变量  $\gamma_i$ ,  $t = 1, \dots, T$ .

不同于传统线性建模方法, Kriging 模型中平稳高斯过程  $Z(x)$  引入相关函数(核函数)度量样本间相关性大小,一方面回归参数  $\beta$  的估计不仅受到输入变量  $x$  的影响,也会受到协方差矩阵的影响,具体体现在参数  $\beta, \sigma^2$  的后验分布中含参数  $\theta$ ; 另一方面,参数  $\theta$  调节非参数部分输入变量

对输出的影响程度,其后验分布受参数  $\beta, \sigma^2$  取值影响. 经过参数逐代抽样,参数  $\theta$  的后验分布随参数  $\beta, \sigma^2$  的更新而更新,进而非参数  $Z(x)$  部分输入对输出的影响程度随  $\theta$  更新而变化.

此外,所提贝叶斯变量筛选过程可退化至参数模型下变量筛选过程. 当 Kriging 模型中参数  $\theta$  取值趋于无穷大时样本间的相关关系为 0, 平稳高斯过程  $Z(x)$  退化为白噪声假设. 在此情况下,样本间协方差矩阵退化为仅包含对角元素的矩阵,输出响应先验分布变为正态分布  $y(x) \sim N(F\beta, \sigma^2 I)$ .

算法 1: 贝叶斯 Kriging 模型构建

输入:  $n$  个训练样本点  $x, y$

步骤 1: 确定全局趋势模型的候选变量集合  $C$ ;

步骤 2: 确定参数  $\sigma^2, \beta, \gamma, \theta$  的先验信息;

步骤 3: 应用贝叶斯分析推断模型参数的后验分布

$$[\beta | \sigma^2, \gamma, \theta, y] \sim N(\bar{A}^{-1} \bar{B} \bar{A}^{-1} \sigma^2)$$

$$[\sigma^2 | \gamma, \theta, y] \sim IG(n/2, (y^T R^{-1} y - \bar{B}^T \bar{A}^{-1} \bar{B})/2)$$

$$p(\theta | \gamma, y) \propto (y^T R^{-1} y - \bar{B}^T \bar{A}^{-1} \bar{B})^{-\frac{n}{2}} |R|^{-\frac{1}{2}} |\bar{A}|^{-\frac{1}{2}} \pi(\theta)$$

$$p(\gamma | \theta, y) \propto (y^T R^{-1} y - \bar{B}^T \bar{A}^{-1} \bar{B})^{-\frac{n}{2}} |D|^{-\frac{1}{2}} |\bar{A}|^{-\frac{1}{2}} \pi(\gamma)$$

步骤 4: 运用 MCMC 依次抽取参数的样本:

- 步骤 4-1: 初始化参数值  $\sigma_{(0)}^2, \beta_{(0)}, \gamma_{(0)}, \theta_{(0)}$ ;
- 步骤 4-2: 从逆伽马分布  $IG$  中抽取  $\sigma_{(t)}^2$ ;
- 步骤 4-3: 从多元正态分布  $N_n$  中抽取  $\beta_{(t)}$ ;
- 步骤 4-4: 从伯努利分布  $B$  中依次抽取  $\gamma_{i(t)}$ ;
- 步骤 4-5: 从  $\theta$  后验分布中抽取  $\theta_{(t)}$ ;

重复步骤 4-2 至 4-5  $T$  次, 得到参数样本序列

$$\{\sigma_{(0)}^2, \beta_{(0)}, \gamma_{(0)}, \theta_{(0)}, \dots, \sigma_{(T)}^2, \beta_{(T)}, \gamma_{(T)}, \theta_{(T)}\}$$

步骤 5: 判断序列是否满足收敛条件, 若满足转步骤 6, 否则转步骤 4;

步骤 6: 利用序列  $\{\gamma_{(1)}, \gamma_{(2)}, \dots, \gamma_{(T)}\}$  估计模型后验概率;

步骤 7: 识别一组后验概率较高的模型;

输出: 一组候选模型  $M = \{M_1, M_2, \dots, M_{10}\}$

### 3.2 基于统计视角的模型有效性检验

在运用贝叶斯方法解决变量选择问题时,通常学者选择后验概率最高的模型作为最终模型,然而 George 和 McCulloch<sup>[38]</sup> 指出最佳模型并不总是表现出最高的后验概率,应保留一组后验概率较高模型作为可能最优模型的候选. 基于此,本研究在考虑变量不确定性的同时将模型结构不确

定性因素也纳入到建模过程中. 以往基于组合模型的研究视角虽通过模型平均方法降低模型不确定性, 但其并不总能显著提高模型的预测精度<sup>[39, 40]</sup>. 考虑到元模型只是对实际过程的近似, 本节从统计检验视角提出基于 Bootstrap 方差检验的贝叶斯 Kriging 模型( Tested Bayesian Kriging with Bootstrap variance, TBK-BV) 对 MCMC 识别出的一组后验概率较高的 Kriging 模型进行有效性检验, 剔除统计上的“无效”模型.

### 3.2.1 基于 Bonferroni 算法修正的多重假设检验

检验贝叶斯 Kriging 模型  $M_\gamma$  有效性的基本思想是通过观察模型在样本测试集上的表现效果, 分析全局趋势模型  $f_s(\mathbf{x})^T \boldsymbol{\beta}_s$  是否能够准确刻画出数据的变化趋势, 对原假设:  $y(\mathbf{x}) \sim N_n(\mathbf{F}_s(\mathbf{x}) \hat{\boldsymbol{\beta}}_s, \hat{\sigma}_s^2 \mathbf{R}(\hat{\boldsymbol{\theta}}_s))$  进行检验以判断模型是否有效. 考虑到高昂的试验成本, 为充分利用训练样本的信息, 采取留一交叉验证法(leave-one-out cross validation, LOO-CV), 对原训练样本集依次剔除样本点  $(\mathbf{x}_i, y_i)$  生成新的样本集合. 在  $n$  个新训练样本集  $(\mathbf{x}_{-i}, y_{-i})$  上分别构建假设检验统计量, 开展候选模型的有效性检验

$$z(\mathbf{x}_i) = \frac{\hat{y}(\mathbf{x}_i) - y(\mathbf{x}_i)}{\hat{s}(\mathbf{x}_i)} \quad (20)$$

值得注意的是, 与一般单个假设检验问题不同, 为避免样本点选取对检验结果的影响, 基于 LOO-CV 的模型检验在每个样本点处均构建了统计量, 因此对于模型  $M_\gamma$  的有效性分析需要同时处理多个假设检验, 对于每个单独的假设检验而言其均可能犯第一类错误, 若不对检验过程进行修正, 最终整体的第一类错误概率将会随着检验次数的增多而升高, 导致模型  $M_\gamma$  在能够有效描述数据变化的情况下被错误拒绝的概率升高. 为控制多重假设检验过程中的整体错误率(family-wise error rate, FWER), 采取 Bonferroni 算法对检验过程进行修正, 将统计量阈值  $z_{\alpha/2}$  调整为  $z_{\alpha/(2n)}$ , 控制 FWER 在显著性水平  $\alpha$  以下.

**定理1** 在显著性水平  $\alpha$  下, 对于  $n$  重假设检验问题, 在 Bonferroni 算法修正情况下 FWER 为

$$FWER = P(V \geq 1) = P\left(\sum_{j \in M_0} I(p_j \leq \alpha/n) \geq 1\right) \leq \alpha \quad (21)$$

其中  $V$  为多重假设检验中错误拒绝原假设的次数,  $p_j$  为第  $j$  个假设  $p$  值,  $M_0$  为原假设为真指标集合.

此外, 在基于 LOO-CV 的 Kriging 模型多重假设检验过程中, 统计量  $z(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  之间统计相关, 不满足独立同分布假设. 考虑到统计量相关情形下的多重假设检验问题的复杂性, 结合模型检验的特点, 对检验过程进行简化处理. 在模型  $M_\gamma$  的有效性检验中, 若存在某个统计量  $z(\mathbf{x}_i) < z_{\alpha/(2n)}$ , 则认为模型  $M_\gamma$  不能准确把握数据的变化趋势, 因此统计量  $z(\mathbf{x}_i)$  检验问题可以转化为检验  $\max |z(\mathbf{x}_i)|$  问题, 通过观察  $\max |z(\mathbf{x}_i)|$  分析模型有效性: 当  $\max |z(\mathbf{x}_i)| < z_{\alpha/(2n)}$ , 即所有样本点处的检验统计量均小于阈值时, 不拒绝原假设; 仅当  $\max |z(\mathbf{x}_i)| \geq z_{\alpha/(2n)}$  时, 拒绝原假设.

### 3.2.2 基于参数 Bootstrap 的方差修正

综合考虑模型预测结果的散度效应, 检验统计量  $z(\mathbf{x}_i)$  中融入了预测方差, 而 Den Hertog<sup>[41]</sup> 指出传统 Kriging 建模估计的预测方差存在低估问题. 为避免有效模型因方差低估而被错误剔除, 采取参数 Bootstrap 对模型预测方差进行修正, 构建更精确的检验统计量.

参数 Bootstrap 是一种利用小样本估计总体的方法, 在不改变数据结构的情况下, 参数 Bootstrap 通过估计总体分布, 伪抽样生成大量新样本, 充分利用有限样本估计总体信息. 为得到参数 Bootstrap 抽样过程中的样本分布, 利用训练样本集  $(\mathbf{x}, \mathbf{y})$  构建 Kriging 模型, 生成总体服从多元正态分布  $N_n(\mathbf{F}_s(\mathbf{x}) \hat{\boldsymbol{\beta}}_s, \hat{\sigma}_s^2 \mathbf{R}(\hat{\boldsymbol{\theta}}_s))$  的样本, 通过从分布中随机抽样获取大量新的子样本, 实现 Bootstrap 预测方差的修正.

- 1) 基于训练集  $(\mathbf{x}, \mathbf{y})$  估计模型超参数  $\{\hat{\boldsymbol{\beta}}_s, \hat{\sigma}_s^2, \hat{\boldsymbol{\theta}}_s\}$ ;
- 2) 从总体分布  $N_n(\mathbf{F}_s(\mathbf{x}) \hat{\boldsymbol{\beta}}_s, \hat{\sigma}_s^2 \mathbf{R}(\hat{\boldsymbol{\theta}}_s))$  中抽取 Bootstrap 样本  $(\mathbf{x}^b, \mathbf{y}^b)$ ;
- 3) 从 Bootstrap 样本  $(\mathbf{x}^b, \mathbf{y}^b)$  中剔除  $(\mathbf{x}_i, y_i)$ ,

基于样本集  $[x_{-i} \ y_{-i}^b]$  估计模型超参数  $\{\hat{\beta}_{s_i}^b, (\hat{\sigma}_s^2)_{-i}^b, \hat{\theta}_{s_i}^b\}$ ;

4) 预测  $x_i$  处的响应值  $\hat{y}^b(x_i)$ ;

5) 重复 2) ~ 4)  $B$  次, 得到  $x_i$  处 Bootstrap 标准

$$s_B(x_i) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{y}^b(x_i) - y^b(x_i))^2};$$

6) 重复 2) ~ 5), 依次剔除样本点, 得到每个样本的 Bootstrap 标准差  $s_B(x_i) \ i=1, 2, \dots, n$ .

基于修正的标准差  $s_B(x_i)$  构建检验统计量  $z_B(x_i)$ , 对候选模型的有效性进行分析. 若对于候选模型  $M_\gamma$  有  $\max |z_B(x_i)| > z_{\alpha/(2n)}$ , 则认为模型  $M_\gamma$  为无效模型, 应将其从候选集中剔除.

值得注意的是, 在大样本情形下基于模型精确度与经济性双重考虑, 所提的算法可以进行灵活调整. 将基于留一交叉验证的统计检验与 Bootstrap 方差修正步骤转变为  $K$  折交叉验证下的分析过程, 确保模型预测的有效性, 兼顾计算资源的有效利用.

### 4 仿真试验

#### 4.1 仿真算例

Borehole 函数作为典型的测试函数, 被众多计算机试验领域的学者研究<sup>[42]</sup>, 为验证第 3 节中 TBK-BV 在变量筛选中的有效性, 本节采取八维 Borehole 函数进行仿真分析

$$y = \frac{2\pi x_3(x_4 - x_6)}{\ln(x_2/x_1) \left[ 1 + \frac{2x_7x_3}{\ln(x_2/x_1)x_1^2x_8} + \frac{x_3}{x_5} \right]} \quad (22)$$

与 Joseph<sup>[18]</sup> 选择试验点的方法相同, 本研究采用拉丁超立方设计在样本空间中选取 27 个设计点作为训练样本. 为避免变量取值范围不一致影响显著变量识别, 在选择初始设计点前对输入变量进行标准化, 将变量取值范围统一限定在  $[0, 1]$  范围内.

#### 4.2 响应建模

在确定初始变量候选集  $C$  时综合考虑了变量一次效应  $x_i$ 、二次效应  $x_i^2$  以及交叉效应  $x_i x_j$ 、 $x_i^2 x_j$ 、 $x_i x_j^2$ 、 $x_i^2 x_j^2$ ,  $i, j = 1, \dots, 8 \ i \neq j$ , 并对候选变量进行了正交多项式编码处理<sup>[37]</sup>. 最终在八维

Borehole 函数问题中, Kriging 全局趋势模型中含有  $128(2d^2)$  个候选变量, 本节探讨如何从  $2^{128}$  个候选模型中筛选出最佳模型.

通过贝叶斯分析得到参数  $[\beta \ \sigma^2 \ \gamma \ \theta]$  后验分布, 应用 MCMC 模拟从参数后验分布中逐代采样, 生成马尔科夫链  $\{\sigma_{(0)}^2, \beta_{(0)}, \gamma_{(0)}, \theta_{(0)}, \dots, \sigma_{(T)}^2, \beta_{(T)}, \gamma_{(T)}, \theta_{(T)}\}$ . 在 100 000 次参数抽样后, 舍弃前 20 000 次燃烧期的样本后, 对剩余的样本每隔四步取样一次, 最终得到 20 000 个参数抽样值<sup>[14]</sup>. 在估计模型后验概率前, 借助自相关图对样本序列的收敛性进行检验. 图 2 列出了部分参数样本序列自相关图.

由图 2 可以观察到随着迭代次数的增加, 样本序列的自相关系数在  $0 \pm 3\sigma$  (图中水平线) 内波动, 由此可见序列满足平稳细致条件. 考虑到模型结构不确定性问题, 保留了十组后验概率较高的变量集合, 分别构建 Kriging 模型作为候选模型集合  $M = \{M_1, M_2, \dots, M_{10}\}$ . 令  $v_k(x)$  为第  $k$  个模型的显著变量集合,  $k = 1, 2, \dots, 10$ , 则与之对应的贝叶斯 Kriging 模型  $M_k$

$$y(x) = v_k(x)^T \beta + Z(x) \quad (23)$$

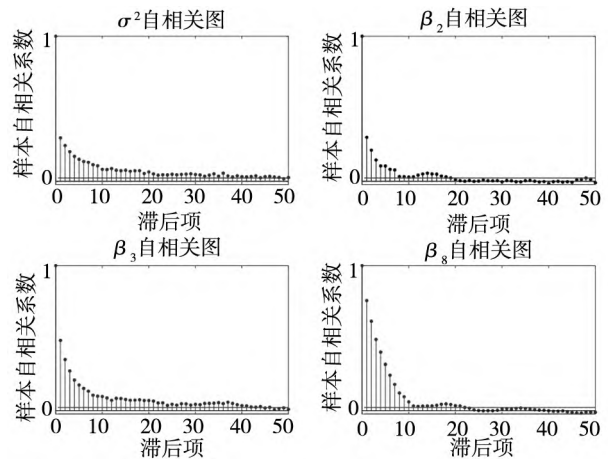


图 2 部分参数样本序列自相关图

Fig. 2 Autocorrelation function plots of several parameter sample sequences

基于统计检验对十组候选模型的有效性进行分析, 通过 LOO-CV 算法构建统计量  $z_{B; k}(x_i)$ , 在 5% 显著性水平下对候选模型逐一进行检验, 剔除其中  $\max |z_{B; k}(x_i)| > z_{\alpha/(2n)}$  的模型, 得到有效模型集合  $M_e$ . 通过拉丁超立方设计选取 1 000 个测

试点, 利用均方根误差  $RMSE$  指标分析模型在测试集上的预测效果, 综合模型的泛化能力选择出最佳模型

$$RMSE = \sqrt{\sum_{i=1}^{n_t} (y_{true}(x_i) - \hat{y}(x_i))^2} \quad (24)$$

### 4.3 方法对比

为验证所提出的 TBK-BV 方法的有效性与鲁棒性, 利用 Borehole 测试函数重复试验 250 次, 同时对比了 OK 模型、UK 模型、基于随机搜索算法的盲 Kriging 模型(the stochastic search blind Kriging, SSBK) 以及基于贝叶斯模型平均的 Kriging 模型(Bayesian model average Kriging, BMAK) 在测试集上的预测效果. 此外, 为体现模型检验中方差修正的必要性, 本研究分析了基于传统方差检验的贝叶斯 Kriging 模型(tested Bayesian Kriging with classic variance, TBK-CV) 结果.

图 3 列出了在 250 次试验中各类方法在测试集上的  $RMSE$  值小提琴图. 通过图 3 可以直观观察到, 在所有模型中未经变量筛选的 UK 模型  $RMSE$  值最高且预测结果的离散程度大. 而在几类变量筛选方法中, 所提的 TBK-BV 模型的  $RMSE$  值最小, 且预测结果集中分布在中位数附近; 相比之下, 采取随机变量搜索算法得到的 SSBK 模型仅考虑了模型的后验概率最高, 而并未从模型有效性角度出发对模型进行分析, 在 250 次重复试验中预测结果较为分散. 此外, 基于

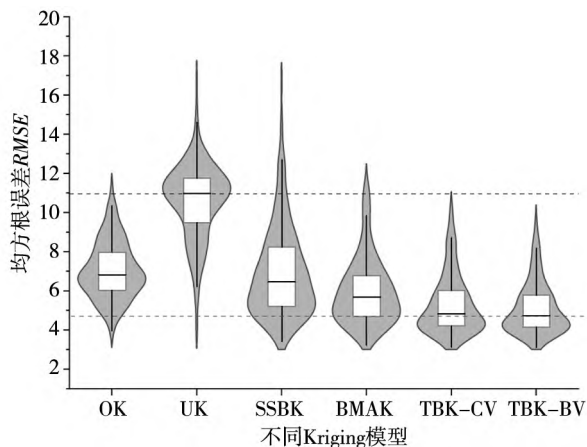


图 3 各种方法在仿真分析中的  $RMSE$  值

Fig. 3 The  $RMSE$  of each method in the simulation analysis

模型平均的 BMAK 模型一方面综合了出现后验

概率较大的几组模型, 表现出了较为稳健的预测效果, 但另一方面无效模型的融入也降低了最终集成的 BMAK 模型预测精度.

考虑到显著性水平与训练样本规模对建模结果的影响, 同时考虑了在 20%、10% 以及 1% 显著性水平下 TBK-BV 方法与其它几类方法在 40(5d)、80(10d) 的训练样本集上的预测效果. 图 4 为显著性水平为 1%、5%、10% 和 20% 下, 各类方法在样本规模为 27、40 以及 80 情况下的模型预测效果.

通过观察图 4 可知, TBK-BV 模型在不同样本规模与显著性水平下均能保持较好的预测效果. UK 模型对样本规模的变动最为敏感, 当样本规模为 27 时, UK 模型预测效果较差, 其  $RMSE$  值约为 TBK-BV 模型  $RMSE$  值的两倍; 仅在样本规模增大至 80 时, UK 模型才表现出与 TBK-BV 模型相当的预测效果. 与 TBK-BV 方法相比, 采取模型组合的 BMAK 方法在样本规模较少时(训练样本数量为 27) 预测效果较差, 只有当训练样本数量不断增加时, BMAK 预测结果  $RMSE$  值折线才逐渐靠近 TBK-BV 的  $RMSE$  值折线. 另外, 值得注意的是, 随着显著性水平的提升, TBK-CV 模型的  $RMSE$  值折线图逐渐上移, 相比之下, TBK-BV 模型的  $RMSE$  值受显著性水平变化影响不大, 因此通过对方差修正可以降低检验统计量对显著性水平变化的敏感程度.

## 5 实例分析

为了进一步验证所提 TBK-BV 模型在实践应用中的有效性, 本节考虑了两类工业情形下的质量设计问题, 一类为变量维数较少且可获得样本数据大于 5d 下的差异性后验概率情形, 一类为变量维数较多而可获得样本数据少于 5d 下的无差异后验概率情形. 探析 TBK-BV 方法在差异性后验概率与无差异后验概率两种情形下的表现效果.

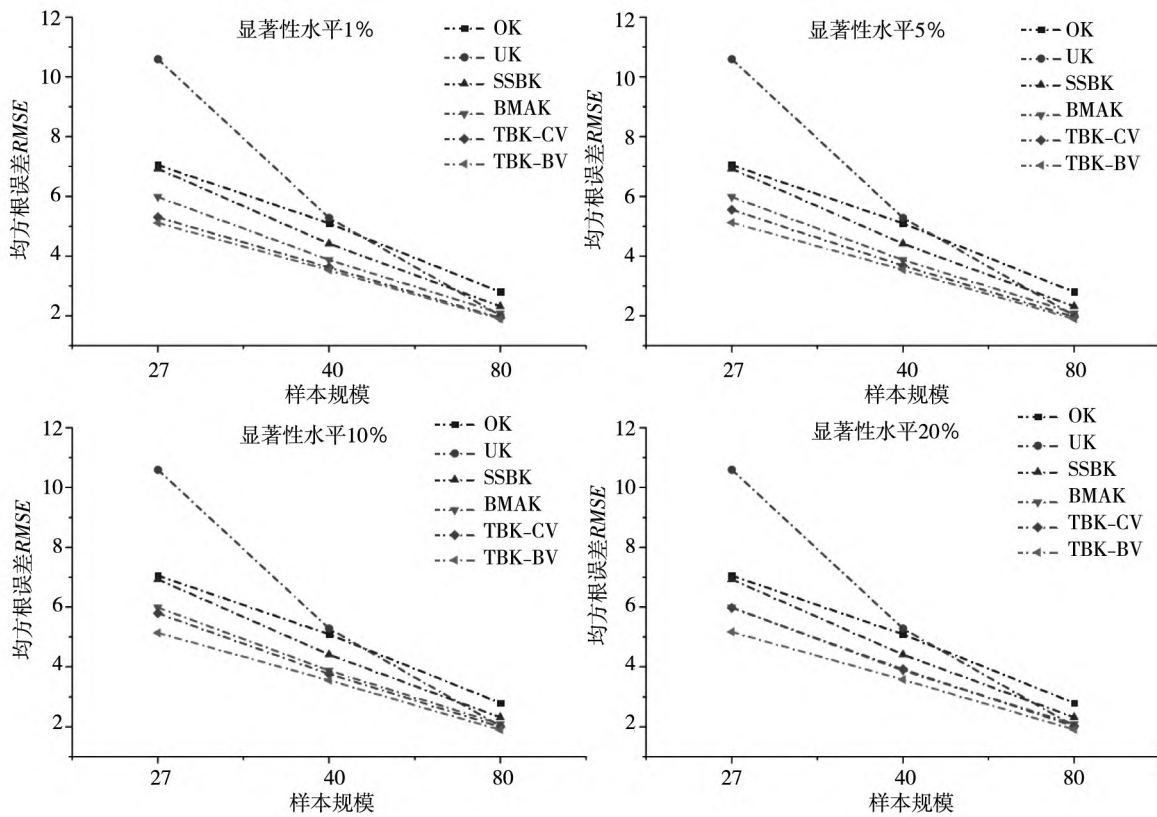


图4 不同显著性水平与样本集大小下各种方法的 RMSE 值

Fig.4 The RMSE of each method under different significance levels and sample sizes

5.1 有差异后验概率验证分析

5.1.1 工业案例: 熔体静电纺丝

增材制造技术因其在生产和设计上的灵活性,在现代制造业中扮演着越来越重要的角色,其中熔体静电纺丝在高分辨率微型3D打印中的应用引起了研究人员的关注.工艺研究表明直径是纺丝过程中至关重要的质量特性,经过显著性分析得到四类影响变量:熔体温度  $x_1$ 、收集板移动速度  $x_2$ 、针头到收集板距离  $x_3$  以及熔体流量  $x_4$ . 本节通过采取中心复合设计得到25组样本,研究样本规模在5d以上的变量筛选与模型检验.

5.1.2 结果分析

综合考虑变量主效应及交互效应对纺丝直径的影响,通过构造贝叶斯Kriging模型对候选变量进行稀疏处理,借助MCMC模拟识别出一组后验概率较高的模型.图5列出了后验概率前十模型中显著变量以及相应模型后验概率大小,其中  $x_{ij}$   $i = 1, 2, 3, 4$  表示变量的一次效应,  $x_{iq}$   $i = 1,$

2, 3, 4 为变量的二次效应.

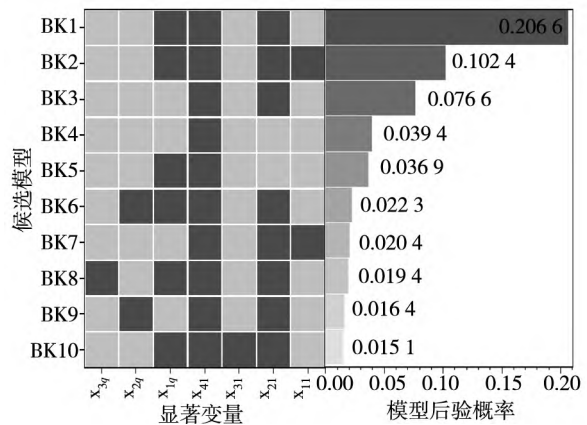


图5 后验概率前十的贝叶斯Kriging模型

Fig.5 Bayesian Kriging models with the top ten posterior probabilities

通过观察图5可以发现在影响熔体静电纺丝直径的因素中,起主导作用的主要为变量的一次效应和二次效应.其中在一次效应里,  $x_{4i}$  被每组模型均识别为显著变量,  $x_{2i}$  被除BK4、BK5之外模型识别为显著变量;在二次效应中,  $x_{1q}$  出现频率较高,由此可知在熔体静电纺丝过程中熔体温

度  $x_1$ 、收集板移动速度  $x_2$  与熔体流量  $x_4$  对直径的影响较大. 此外, 通过观察图 5 中 BK1 ~ BK10 模型后验概率分布情况可以发现, MCMC 模拟中出现频率最高的模型 BK1 后验概率高达 0.206 6, 显著高于其它模型的后验概率.

表 1 中列出了各个模型在传统方差与 Bootstrap 方差下的检验结果与预测表现. 在 5% 显著性水平下经传统方差与修正方差检验得到有效模型分别为 {BK3, BK4, BK7, BK9} 与 {BK1, BK3, BK4, BK5, BK6, BK7, BK8, BK9, BK10}. 相对于传统预测方差, Bootstrap 方差由于对低估方差进行了修正, 减少了候选模型被误判为无效模型的概率. 在有效模型集合上综合考察候选模型的泛化能力, 最终 TBK-BV 方法识别出 Kriging 全局趋势模型中显著变量为  $x_{2l}, x_{4l}, x_{1q}$ .

表 1 候选模型的检验结果和预测表现

Table 1 The test results and predictive performance of candidate models

模型	显著变量集合	max z	max z <sub>B</sub>	RMSE
BK1	$x_{2l}, x_{4l}, x_{1q}$	7.025	2.787	0.287
BK2	$x_{1l}, x_{2l}, x_{4l}, x_{1q}$	4.071	3.360	0.420
BK3	$x_{2l}, x_{4l}$	2.793	2.401	0.705
BK4	$x_{4l}$	2.873	1.844	0.625
BK5	$x_{4l}, x_{1q}$	3.217	1.764	0.634
BK6	$x_{2l}, x_{4l}, x_{1q}, x_{2q}$	3.515	2.881	0.418
BK7	$x_{1l}, x_{2l}, x_{4l}$	2.535	2.051	1.058
BK8	$x_{2l}, x_{4l}, x_{1q}, x_{3q}$	6.990	2.580	0.393
BK9	$x_{2l}, x_{4l}, x_{2q}$	2.707	2.770	0.776
BK10	$x_{2l}, x_{3l}, x_{4l}, x_{1q}$	3.143	2.853	0.306
TBK-BV	$x_{2l}, x_{4l}, x_{1q}$	7.025	2.787	0.287
TBK-CV	$x_{4l}$	2.873	1.844	0.625
SSBK	$x_{2l}, x_{4l}, x_{1q}$	7.025	2.787	0.287
UK	$x_{1l}, x_{2l}, x_{3l}, x_{4l}$	2.445	2.113	1.003
OK	—	6.496	2.877	0.806
BMAK	—	7.025	2.722	0.323

通过观察表 1 各个方法在测试集上的 RMSE 值可知, 所提的 TBK-BV 模型与 SSBK 模型识别出的显著变量一致. 与未经变量筛选的 OK 和 UK 模型相比, TBK-BV 方法的 RMSE 分别降低了 64.39% 和 71.39%, 由此可见经变量筛选剔除冗余项能够有效提高模型预测精度. TBK-CV 方法

由于低估的方差使其错误拒绝了后验概率最高且表现最好的模型 BK1, 最终表现较 TBK-BV 方法差. 而 BMAK 方法与所提 TBK-BV 方法相比, 基于模型组合的思想增加预测工作量, 同时无效模型的融入导致模型精度降低, 最终其 RMSE 相比于 TBK-BV 方法的 RMSE 高 12.54%.

## 5.2 无差异后验概率验证分析

### 5.2.1 工业案例: 活塞拍击噪声

汽车活塞的次级运动引起的引擎噪声是导致顾客满意度水平下降的原因之一, 工艺研究表明影响活塞噪音大小的因素主要有六个: 活塞与缸套间隙  $x_1$ 、压力峰值位置  $x_2$ 、活塞裙长度  $x_3$ 、活塞裙部型线  $x_4$ 、活塞裙部轮廓  $x_5$  以及活塞销偏置  $x_6$ . 综合考虑到因素主效应及交互效应, 候选变量多达 72 项, 而运行一次试验成本较高, 因此选取 12 组数据作为训练数据<sup>[18]</sup>, 研究样本规模在 5d 以下的模型变量筛选与统计检验.

### 5.2.2 结果分析

经过贝叶斯变量筛选, 图 6 列出了后验概率前十模型中显著变量分布及相应后验概率. 通过观察图 6 可知, 在活塞噪声中影响输出的变量主要为  $x_{1l}, x_{6l}, x_{2q}, x_{4q}, x_{6q}, x_{1l}x_{5l}, x_{2q}x_{3l}, x_{1l}x_{5q}, x_{3l}x_{4q}, x_{2q}x_{3q}$ . 相对于筛选前 72 维变量, 所提方法显著减少了变量. 其中  $x_1$  一次效应与  $x_4$  二次效应被十组模型中均识别为显著变量, 且交互项  $x_{1l}x_{5l}$  被较多模型识别为显著变量, 由此可见活塞与缸套间隙、活塞裙部型线及轮廓对活塞噪声的影响较大.

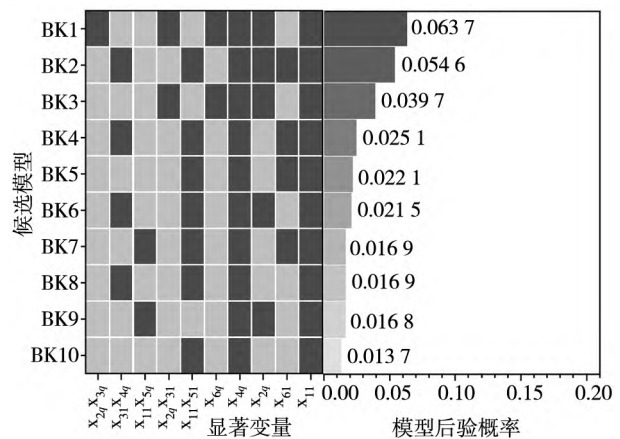


图 6 后验概率前十的贝叶斯 Kriging 模型

Fig. 6 Bayesian Kriging models with the top ten posterior probabilities

观察图 6 中各个模型的后验概率可以发现,

与 5.1 小节熔体静电纺丝案例建模分析结果不同,在高维样本量较少的活塞噪音案例中,可获得的样本信息较少,经过 MCMC 模拟后得到的十组候选模型后验概率差异性小,主要分布在 0.06 ~ 0.01 附近,建模过程中模型的不确定性较高。在此情形下,为避免出现后验概率差异性不大导致最佳模型未能被最高后验概率识别的情况,需要考虑模型结构不确定性影响。

图 7 反映了 BK1 ~ BK10 模型在传统预测方差下检验统计量(蓝色散点)与基于 Bootstrap 修正的预测方差下检验统计量(红色散点)的分布情况。通过观察图 7 可知,在 5% 的显著性水平下,当采取传统方差对模型有效性展开检验时,十组模型的检验统计量值均落在拒绝域。而经过 Bootstrap 方差修正后,模型的检验统计量的值均出现不同程度的降低,最终得到后验概率最高的 BK1 模型以及其它四组模型均通过检验,进一步体现了在模型检验过程中对模型预测方差进行修正的必要性。综合统计检验与模型泛化能力分析,最终识别出显著变量为  $x_{1l}, x_{6l}, x_{4q}, x_{1l}, x_{5l}$ 。

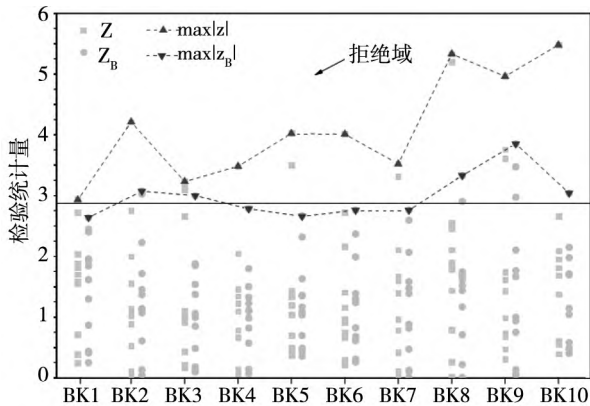


图 7 贝叶斯 Kriging 模型统计检验

Fig. 7 Statistical test for Bayesian Kriging models

表 2 各种方法在测试集上的预测结果

Table 2 Predictive performance of each method on the test set

模型	显著变量	RMSE
OK	—	1.364
UK	$x_{1l}, x_{2l}, x_{3l}, x_{4l}, x_{5l}, x_{6l}$	1.511
SSBK	$x_{1l}, x_{2q}, x_{4q}, x_{6q}, x_{2q}, x_{3l}, x_{2q}, x_{3q}$	1.823
BMAK	—	1.032
TBK-BV	$x_{1l}, x_{6l}, x_{4q}, x_{1l}, x_{5l}$	0.842

表 2 列出了 OK、UK、BMAK 和 SSBK 以及

TBK-BV 模型在测试集上的 RMSE 值。通过观察表 2 可知,提出的 TBK-BV 方法在所列方法中表现最优。其中 BMAK 方法基于模型平均的思想综合候选模型预测效果,虽表现优于传统 OK、UK 以及 SSBK 方法,但由于在模型平均时融入了无效模型,其预测结果 RMSE 值较 TBK-BV 的 RMSE 值高 22.57%。此外,相比于仅考虑模型后验概率,未从统计角度分析模型有效性的 SSBK 模型,所提 TBK-BV 方法的 RMSE 降低了 53.82%。由此可见,在高维且样本量较少情形下,经过贝叶斯分析得到的一组候选模型后验概率差异性小,模型的不确定性使得最佳模型不一定能够由最高后验概率识别,此时有必要保留一组后验概率较高的模型并对模型结构不确定性进行分析。

### 5.3 讨论

实例分析结果表明,所提方法无论在有、无差异后验概率验证分析中均能获得较好的预测结果。在候选变量多而样本信息较少情形下,模型不确定性较高,经过贝叶斯分析和 MCMC 模拟得到的候选模型后验概率相近。此时,若采取随机变量搜索算法直接选择后验概率最高的模型,则可能受模型不确定性影响导致所选模型非最佳模型。若采取基于模型平均 BMAK 方法,虽能通过组合建模技术提高模型的预测结果稳定性,但也会因融入表现较差的模型而导致模型预测精度降低。而所提 TBK-BV 方法充分考虑了模型不确定性,保留一组“有潜力”的候选模型,综合模型有效性与泛化能力得到预测效果较优的模型。相比于前一类情形,在候选变量维度较低且样本较多情形下,建模过程中的不确定性较小,后验概率最高的模型与后验概率次高的模型概率相差显著,经过模型有效性检验,最终 TBK-BV 模型与利用随机变量搜索算法得到的最佳模型 SSBK 识别出的变量结果一致。

## 6 结束语

产品复杂程度的提升与数字技术的发展推动着基于计算机试验的代理模型在质量设计领域的

应用. 本研究针对质量设计中关键变量选择问题, 综合考虑质量变量不确定性与模型结构不确定性, 在贝叶斯框架下融合因子效应原则与假设检验思想, 提出基于统计检验的贝叶斯 Kriging 模型, 以实现高效的变量筛选, 为后续质量改进与优化奠定基础. 为验证所提方法在质量设计过程中的有效性和稳健性, 采取 8 维 Borehole 函数研究在不同的样本规模与不同显著性水平下模型的预测效果, 并通过不同工业情景下质量建模, 考察了所提方法在差异性后验概率与无差异后验概率两种情形下的预测效果. 本研究的主要研究结果如下.

(1) 在影响因子与质量特性间关系未知的情况下, 结合因子效应原则和贝叶斯方法显著降低了 Kriging 全局趋势模型中的变量维度, 同时从统计角度对筛选后模型的有效性展开检验, 综合分析了变量不确定性与模型结构不确定性的影响;

(2) 相比于现有的变量筛选方法, 所提 TBK-BV 方法在不同的样本规模与不同的显著性水平

下均表现出较优的预测效果, 即使在可获取样本信息较少的情形下仍然能够实现高效的变量筛选;

(3) 由于变量不确定性与模型结构不确定性广泛存在于其它代理建模中, 如支持向量机模型<sup>[43]</sup>等, 因此所提思想同样适用于其他模型, 本研究可进一步丰富和发展代理建模在质量设计的研究内容.

本研究在研究质量设计元建模时假设在特定的输入工艺参数下输出响应唯一确定, 后续将进一步在随机 Kriging 模型框架下, 综合分析在内外噪声影响下稀疏元建模问题. 同时, 模型序贯加点能够有效提高预测精度, 因此如何将 Kriging 模型变量筛选与序贯优化设计结合有待未来深入研究. 此外, 受建模时间复杂度与空间复杂度的限制, Kriging 模型主要用于样本规模较小情形下的建模问题, 而随着信息技术的发展, 大规模样本给质量建模与设计带来了挑战, 因此, 如何在高维样本情形下进行 Kriging 建模与变量筛选也是未来有待深入研究的方向之一.

#### 参考文献:

- [1] 刘心报, 胡俊迎, 陆少军, 等. 新一代信息技术环境下的全生命周期质量管理[J]. 管理科学学报, 2022, 25(7): 2-11.  
Liu Xinbao, Hu Junying, Lu Shaojun, et al. The entire life cycle quality management in the new generation of information technology environment[J]. Journal of Management Sciences in China, 2022, 25(7): 2-11. (in Chinese)
- [2] 张玺, 宋洁, 侍乐媛, 等. 新一代信息技术环境下的高端装备数字化制造协同[J]. 管理世界, 2023, 39(1): 190-204.  
Zhang Xi, Song Jie, Shi Leyuan, et al. Collaboration for digital manufacturing of high-end equipment in the era of new-generation information technology[J]. Management World, 2023, 39(1): 190-204. (in Chinese)
- [3] 蒋炜, 李四杰, 黄文坡, 等. 物联网大数据与产品全生命周期质量管理[M]. 北京: 科学出版社, 2021.  
Jiang Wei, Li Sijie, Huang Wenpo, et al. Internet of Things Big Data and Product Life Cycle Quality Management[M]. Beijing: Science Press, 2021. (in Chinese)
- [4] Fisher R A. The Design of Experiments[M]. Edinburgh: Oliver and Boyd, 1937.
- [5] Wu C F J. Post-fisherian experimentation: From physical to virtual[J]. Journal of the American Statistical Association, 2015, 110(510): 612-620.
- [6] Wang S, Ng S H, Haskell W B. A multilevel simulation optimization approach for quantile functions[J]. INFORMS Journal on Computing, 2022, 34(1): 569-585.
- [7] 欧阳林寒, 陶宝平, 马妍. 基于选择性集成核高斯过程模型的质量预测研究[J]. 中国管理科学, 2023, 31(3): 69-80.

- Ouyang Linhan , Tao Baoping , Ma Yan. Research on quality prediction based on Gaussian process model with selective ensemble kernel [J]. *Chinese Journal of Management Science* , 2023 , 31( 3 ) : 69 – 80. ( in Chinese)
- [8] 闫伟 , 何桢 , 李岸达. 基于 CEM-IG 算法的复杂产品关键质量特性识别 [J]. *系统工程理论与实践* , 2014 , 34( 5 ) : 1230 – 1236.
- Yan Wei , He Zhen , Li Anda. Identification of critical-to-quality characteristics for complex products using CEM-IG algorithm [J]. *Systems Engineering: Theory & Practice* , 2014 , 34( 5 ) : 1230 – 1236. ( in Chinese)
- [9] Shi W , Chen X , Shang J. An efficient Morris method-based framework for simulation factor screening [J]. *INFORMS Journal on Computing* , 2019 , 31( 4 ) : 745 – 770.
- [10] Wang Z , Wang H , Wang S , et al. Linear mixed-effects model for longitudinal complex data with diversified characteristics [J]. *Journal of Management Science and Engineering* , 2020 , 5( 2 ) : 105 – 124.
- [11] Ouyang L , Zhu S , Ye K , et al. Robust Bayesian hierarchical modeling and inference using scale mixtures of normal distributions [J]. *IIE Transactions* , 2022 , 54( 7 ) : 659 – 671.
- [12] Han M , Liu X , Huang M , et al. Integrated parameter and tolerance optimization of a centrifugal compressor based on a complex simulator [J]. *Journal of Quality Technology* , 2020 , 52( 4 ) : 404 – 421.
- [13] 崔庆安 , 崔楠. 试验设计及参数优化的 LS-SVR 显著性因子筛选 [J]. *管理科学学报* , 2023 , 26( 12 ) : 42 – 61.
- Cui Qing'an , Cui Nan. Significant factors screening of LS-SVR for experimental design and parameters optimization [J]. *Journal of Management Sciences in China* , 2023 , 26( 12 ) : 42 – 61. ( in Chinese)
- [14] 汪建均 , 屠雅楠 , 马义中. 结合 SUR 与因子效应原则的多响应质量设计 [J]. *管理科学学报* , 2020 , 23( 12 ) : 12 – 29.
- Wang Jianjun , Tu Yanan , Ma Yizhong. Multi-response quality design integrating SUR models with factorial effect principles [J]. *Journal of Management Sciences in China* , 2020 , 23( 12 ) : 12 – 29. ( in Chinese)
- [15] Montgomery D C. *Design and Analysis of Experiments* [M]. New York: John Wiley & Sons , 2013.
- [16] 施文 , 刘志学 , 刘丹. 零部件循环取货越库物流系统关键因子研究——仿真实验设计与分析方法 [J]. *管理科学学报* , 2013 , 16( 6 ) : 22 – 34.
- Shi Wen , Liu Zhixue , Liu Dan. Key factors of parts logistics for milk-run and cross-docking system: Design and analysis of simulation experiments method [J]. *Journal of Management Sciences in China* , 2013 , 16( 6 ) : 22 – 34. ( in Chinese)
- [17] Chen Z , Han S. Comparison of dimension reduction methods for DEA under big data via Monte Carlo simulation [J]. *Journal of Management Science and Engineering* , 2021 , 6( 4 ) : 363 – 376.
- [18] Joseph V R , Hung Y , Sudjianto A. Blind Kriging: A new method for developing metamodels [J]. *Journal of Mechanical Design* , 2008 , 130( 3 ) : 031102.
- [19] Huang H , Lin D K J , Liu M Q , et al. Variable selection for Kriging in computer experiments [J]. *Journal of Quality Technology* , 2020 , 52( 1 ) : 40 – 53.
- [20] Deng H , Shao W , Ma Y , et al. Bayesian metamodeling for computer experiments using the Gaussian Kriging models [J]. *Quality and Reliability Engineering International* , 2012 , 28( 4 ) : 455 – 466.
- [21] Hung Y. Penalized blind Kriging in computer experiments [J]. *Statistica Sinica* , 2011 , 21( 3 ) : 1171 – 1190.
- [22] Zhang Y , Yao W , Chen X , et al. A penalized blind likelihood Kriging method for surrogate modeling [J]. *Structural and Multidisciplinary Optimization* , 2020 , 61( 2 ) : 457 – 474.
- [23] Park I. Lasso Kriging for efficiently selecting a global trend model [J]. *Structural and Multidisciplinary Optimization* , 2021 , 64( 3 ) : 1527 – 1543.
- [24] Shao W , Deng H , Ouyang L , et al. A type-II maximum-likelihood approach to Gaussian scale mixture-based sparse regression Kriging [J]. *Computers & Industrial Engineering* , 2022 , 168: 108028.
- [25] 林成龙 , 马义中 , 肖甜丽 , 等. 基于 PLS 方法和 Kriging 模型的多目标有效全局优化方法及应用 [J]. *系统工程理论与实践* , 2021 , 41( 7 ) : 1855 – 1867.

- Lin Chenglong , Ma Yizhong , Xiao Tianli , et al. Multi-objective efficient global optimization method and its application based on PLS method and Kriging model [J]. *Systems Engineering: Theory & Practice* , 2021 , 41( 7) : 1855 – 1867. ( in Chinese)
- [26] Park C , Borth D J , Wilson N S , et al. Variable selection for Gaussian process regression through a sparse projection [J]. *IIE Transactions* , 2022 , 54( 7) : 699 – 712.
- [27] Kleijnen J P C , Van Beers W C M. Statistical tests for cross-validation of Kriging models [J]. *INFORMS Journal on Computing* , 2022 , 34( 1) : 607 – 621.
- [28] Demay C , Iooss B , Le Gratiet L , et al. Model selection based on validation criteria for Gaussian process regression: An application with highlights on the predictive variance [J]. *Quality and Reliability Engineering International* , 2022 , 38( 3) : 1482 – 1500.
- [29] Ba S , Joseph V R. Composite Gaussian process models for emulating expensive functions [J]. *The Annals of Applied Statistics* , 2012 , 6( 4) : 1838 – 1860.
- [30] 肖甜丽 , 马义中 , 林成龙. 面向质量设计的 Kriging 组合建模技术 [J]. *计算机集成制造系统* , 2021 , 27( 7) : 2023 – 2034.
- Xiao Tianli , Ma Yizhong , Lin Chenglong. Ensemble Kriging modeling technique for quality design [J]. *Computer Integrated Manufacturing Systems* , 2021 , 27( 7) : 2023 – 2034. ( in Chinese)
- [31] 万良琪 , 欧阳林寒. 基于 0-1 规划模型筛选策略的 Kriging 组合模型及可靠性优化设计 [J]. *计算机集成制造系统* , 2022 , 28( 7) : 2162 – 2168.
- Wan Liangqi , Ouyang Linhan. Kriging ensemble model based on 0-1 programming model selection strategy for reliability-based design optimization [J]. *Computer Integrated Manufacturing Systems* , 2022 , 28( 7) : 2162 – 2168. ( in Chinese)
- [32] Zhang J , Taflanidis A A. Bayesian model averaging for Kriging regression structure selection [J]. *Probabilistic Engineering Mechanics* , 2019 , 56: 58 – 70.
- [33] Lophaven S N , Nielsen H B , Søndergaard J. DACE: A Matlab Kriging Toolbox [M]. Lyngby: Informatics and Mathematical Modelling , Technical University of Denmark , 2002.
- [34] Joseph V R , Delaney J D. Functionally induced priors for the analysis of experiments [J]. *Technometrics* , 2007 , 49( 1) : 1 – 11.
- [35] Chen Z , Mak S , Wu C F J. A hierarchical expected improvement method for Bayesian optimization [J]. *Journal of the American Statistical Association* , 2024 , 119( 546) : 1619 – 1632.
- [36] Chipman H , Hamada M , Wu C F J. A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing [J]. *Technometrics* , 1997 , 39( 4) : 372 – 381.
- [37] Wu C F J , Hamada M S. *Experiments: Planning , Analysis , and Optimization* [M]. New York: John Wiley & Sons , 2011.
- [38] George E I , McCulloch R E. Variable selection via Gibbs sampling [J]. *Journal of the American Statistical Association* , 1993 , 88( 423) : 881 – 889.
- [39] Yuan Z , Yang Y. Combining linear regression models: When and how? [J]. *Journal of the American Statistical Association* , 2005 , 100( 472) : 1202 – 1214.
- [40] Yang Y. Regression with multiple candidate models: Selecting or mixing? [J]. *Statistica Sinica* , 2003 , 13( 3) : 783 – 809.
- [41] Den Hertog D , Kleijnen J P C , Siem A Y D. The correct Kriging variance estimated by bootstrapping [J]. *Journal of the Operational Research Society* , 2006 , 57( 4) : 400 – 409.
- [42] Sung C L , Wang W , Plumlee M , et al. Multiresolution functional ANOVA for large-scale , many-input computer experiments [J]. *Journal of the American Statistical Association* , 2020 , 115( 530) : 908 – 919.
- [43] 崔庆安. 面向多极值质量特性的过程参数全局优化研究 [J]. *管理科学学报* , 2012 , 15( 9) : 46 – 57.
- Cui Qingan. Global optimization of process parameters for multi-extreme quality characteristics [J]. *Journal of Management*

Sciences in China ,2012 ,15(9) : 46 -57. ( in Chinese)

## Metamodeling in computer experiments for quality design: A statistical test perspective

*OUYANG Lin-han*<sup>1,2</sup> , *TAO Bao-ping*<sup>1</sup> , *HE Zhen*<sup>3\*</sup>

1. College of Economics and Management , Nanjing University of Aeronautics and Astronautics , Nanjing 211106 , China;
2. Humanities and Social Sciences Laboratory of Jiangsu Province for Digital Intelligence Management and Low-carbon Operations for Manufacturing System , Nanjing University of Aeronautics and Astronautics , Nanjing 211106 , China;
3. College of Management and Economics , Tianjin University , Tianjin 300072 , China

**Abstract:** This paper presents a novel Bayesian Kriging model for quality design which tackles both variable uncertainty and model structure uncertainty in the metamodeling process , thus providing a robust foundation for quality improvement. Within the Bayesian hierarchical framework , significant variables in the global trend model of Kriging are effectively identified , and the validity of the candidate models is rigorously assessed through statistical tests. Initially , factor effect principles are integrated into the prior distributions of parameters to clarify their relationships , thus significantly reducing the dimensionality of the candidate space. Subsequently , Markov Chain Monte Carlo simulations are employed to estimate the posterior probabilities of the models , identifying Kriging models with sparse global trend structures. The validity of candidate models is then analyzed through multiple hypothesis testing , with corrections applied to the underestimated prediction variance. The final model is selected based on a comprehensive assessment of both its validity and generalization capability. The simulation results indicate that the proposed method performs satisfactorily across different sample sizes and significance levels. Additionally , the results of the case studies under two industrial scenarios demonstrate that the proposed method effectively identifies significant variables under both differential and non-differential posterior probability conditions.

**Key words:** quality design; Kriging model; Bayesian hierarchical model; variable selection; statistical test