

基于SCAD方法的烘丝段水分控制参数筛选及应用

许磊

云南贝叶思大数据技术服务有限公司, 云南 昆明

收稿日期: 2025年2月18日; 录用日期: 2025年3月11日; 发布日期: 2025年3月25日

摘要

为实现对制丝过程烘丝段出料水分的精准控制, 科学、客观地识别并筛选影响水分的关键工艺参数。选取以某牌号卷烟的制丝过程全批次稳态数据为研究对象, 在传统逐步回归方法的基础上, 引入三种通过添加惩罚项来压缩变量系数的Lasso族方法, 即: Lasso方法、适应性Lasso方法和SCAD方法, 分别构建以烘丝段出料水分为因变量的变量选择模型, 并采用AIC、BIC和MSE三种评价指标对模型进行比较, 最后依据最优模型进行关键工艺参数的筛选及重要性排序。结果表明: ① Lasso族方法在模型拟合优度、预测精度和复杂度控制方面均显著优于传统逐步回归方法, 其中SCAD方法的综合性能表现最优; ② 基于SCAD方法确定了烘丝段的4个关键工艺参数, 按其重要性排序依次为: II区筒壁温度、膨胀单元蒸汽体积流量、切叶丝含水率和工艺气速度。

关键词

烘丝段, 水分控制, 变量选择, 模型比较, SCAD方法

Selection and Application of Moisture Control Parameters in the Drying Section Based on the SCAD Method

Lei Xu

Yunnan Bayes Big Data Technology Service Co., Kunming Yunnan

Received: Feb. 18th, 2025; accepted: Mar. 11th, 2025; published: Mar. 25th, 2025

Abstract

To achieve precise control over the moisture content of the discharge in the drying section of the

tobacco primary processing, it is essential to scientifically and objectively identify and screen the key process parameters affecting moisture. This study selects steady state data from the entire batch of a specific brand of cigarette production process as the research object. Building on the traditional stepwise regression method, three Lasso family methods—Lasso, Adaptive Lasso, and SCAD are introduced, which compress variable coefficients by adding penalty terms. Variable selection models are constructed with the moisture content of the drying section discharge as the dependent variable. The models are compared using three evaluation metrics: AIC, BIC, and MSE. Finally, the optimal model is used to screen and rank the importance of key process parameters. The results show that: ① The Lasso-family methods significantly outperform the traditional stepwise regression method in terms of model goodness-of-fit, prediction accuracy, and complexity control, with the SCAD method demonstrating the best overall performance; ② Based on the SCAD method, four key process parameters for the drying section are identified, ranked in order of importance as follows: Zone II wall temperature, expansion unit steam volumetric flow rate, cut tobacco moisture content, and process air velocity.

Keywords

Drying Section, Moisture Control, Parameter Selection, Model Comparison, SCAD Method

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在卷烟生产过程中，制丝环节是决定产品质量和生产效率的关键环节，其生产效能和成品烟丝质量直接影响烟草企业的经济效益[1]。其中，烘丝段作为制丝过程的核心生产工序，其出料水分的稳定性对烟丝的物理特性和内在特性具有显著影响[2]。然而，烘丝段涉及的设备参数与工艺参数种类繁多，且各参数之间往往存在复杂的交互作用，在实际生产过程中全面精准地调控所有参数并不现实。因此，筛选关键工艺参数并对其进行重点调控，是实现出料水分稳定控制、满足生产工艺标准的必要途径。在此背景下，采用科学合理的变量选择方法对烘丝段工艺参数进行筛选，具有重要的理论意义和实践价值。

在烟草相关问题的研究中，变量选择是构建预测模型和优化加工参数的关键步骤。传统方法通常采用多元逐步回归和信息准则进行变量选择，并在相关研究中得到了广泛应用。钟文焱等[2]通过神经网络算法和多元回归分析方法建立含水率预测模型；谭超等[3]通过偏最小二乘和后向区间选择压缩变量数的方法分析制丝工艺烘丝工序后加香工序前烟草糖含量；卫盼盼等[4]通过逐步回归、偏相关和通径分析得到各级打辊速度和风分风速显著相关的物理指标；王明锋等[5]通过逐步回归的方法选择对感官舒适性指标有显著影响的化学成分；李晋明等[6]采用逐步回归分析法从萃取物中筛选出关键致香成分。然而，尽管上述传统变量选择方法在烟草及相关领域展现了一定的实用性，但其固有的局限性也不容忽视。Breiman [7]指出，传统变量选择方法往往缺乏稳定性，即在不同样本或不同条件下，所选出的变量可能会有较大差异；Fan [8]进一步指出，传统方法的计算过程容易受到随机误差的影响，且在处理大型分类或回归问题时，其计算复杂度往往超出计算机的处理能力。针对传统变量选择方法的缺陷，Tibshirani [9]提出了 Lasso (Least Absolute Shrinkage and Selection Operator)方法，即通过在最小二乘估计中引入 L1 惩罚项，同时实现变量筛选和参数估计；Efron [10]提出的最小角回归算法有效解决了 Lasso 方法的计算效率问题；Lasso 方法的主要优势在于其计算过程具有顺序性和连续性，且能够有效处理变量间的多重共线性问题。然而，该方法也存在明显不足：其对所有变量施加相同的惩罚，导致估计量存在偏差，且不满足

Oracle 性质(即变量选择的稀疏性、连续性和无偏性)。为了弥补 Lasso 方法的缺陷, Zou [11]提出了适应性 Lasso (Adaptive Lasso)方法, 通过引入惩罚权重改进了变量选择的准确性; Fan 和 Li [8]则针对 Lasso 方法中系数过度压缩的问题, 提出了 SCAD (Smoothly Clipped Absolute Deviation)惩罚方法, 进一步提升了变量选择的性能。鉴于此, 本文以制丝过程烘丝段为研究对象, 在传统逐步回归方法的基础上, 引入三种 Lasso 族方法开展出料水分的关键影响参数筛选研究, 旨在为制丝过程水分控制的精准性提供新的思路和方法参考。

2. 数据与方法

2.1. 数据来源及预处理

研究数据来源于某卷烟厂 MES 系统在 2024 年 1 月至 2024 年 12 月期间采集的牌号 A 卷烟制丝过程全批次数据, 共计 219 批, 数据采集频次为每 6 秒采集 1 次。由于卷烟加工过程在起始和结束阶段可能存在不稳定状态, 导致采集到的数据中包含停机断料、非稳态等异常情况, 因此对原始数据进行了系统性预处理, 具体步骤如下: 1) 停机断料批次剔除, 设定判定条件为烘丝段工艺流量降至 0 kg/h 且持续时间超过 90 秒, 符合该条件的批次被判定为停机断料批次并予以剔除, 共剔除了 21 个异常批次; 2) 稳态数据截取, 依据《中式卷烟制丝生产线创新效果评价测试大纲》中的稳态截取规则, 去除各批次起始阶段和结束阶段的非稳态数据, 保留中间稳定生产时段的数据, 此步骤旨在消除加工过程中因设备启停或工况波动导致的非稳态干扰; 3) 针对制丝滚筒类设备混合加工特性导致的工艺参数与出料水分实时数据无法一一对应的问题, 采用物料停留时间测定方法进行时序对齐。通过实测制丝生产过程标识物(纯白卷烟纸)在烘丝段工序的停留时间, 确定物料停留时间为 8 分钟。基于此, 以 8 分钟为时间间隔对稳态数据进行分组, 并计算每组数据的均值, 最终形成具有时序对应关系的稳态数据样本。

2.2. 参数类型及变量选取

在制丝生产的复杂流程中, 烘丝工序段的控制参数最多, 可分为设备参数与工艺参数两大类。工艺参数是操作人员通过精确设定与动态调节来控制出料水分的关键手段, 直接影响最终产品的质量。而设备参数则通过自动化系统进行反馈控制, 无法人为直接干预, 其作用在于确保工艺参数在加工过程中的稳定性。在样本数据中, 部分参数在加工过程中保持固定不变, 不纳入本文的研究范围。烘丝段参数类型和变量选取如表 1 所示。

Table 1. Process parameters and equipment parameters of the tobacco drying section
表 1. 烘丝段工艺参数和设备参数

参数类别	变量	参数名称	备注
	Y	出料水分	因变量
设备参数	X ₁	SX 蒸汽阀门开度	自变量
	X ₂	物料累计量	
	X ₃	排潮阀门开度	
	X ₄	循环风阀门开度	
	X ₅	热交换器薄膜阀阀门开度	
	X ₆	筒壁二区蒸汽阀门开度	
	X ₇	切叶丝含水率	

续表

工艺参数	X_8	工艺流量
	X_9	膨胀单元蒸汽体积流量
	X_{10}	膨胀单元蒸汽质量流量
	X_{11}	负压
	X_{12}	工艺气速度
	X_{13}	I 区筒壁温度
	X_{14}	II 区筒壁温度
	X_{15}	热风温度

根据烘丝机的控制原理，以上设备参数和工艺参数可能产生三类交互效应，即：设备参数之间的交互效应、工艺参数之间的交互效应、设备参数与工艺参数之间的交互效应，具体结果如表 2 所示。

Table 2. The interaction effects between parameters in the tobacco drying section
表 2. 烘丝段参数间的交互效应

交互效应类别	自变量
设备参数间的交互效应	$X_3 X_4$
工艺参数间的交互效应	$X_9 X_{10}$
设备参数与工艺参数间的交互效应	$X_1 X_9$ 、 $X_1 X_{10}$ 、 $X_1 X_9 X_{10}$ 、 $X_3 X_{11}$ 、 $X_4 X_{11}$ 、 $X_3 X_4 X_{11}$ 、 $X_5 X_8$ 、 $X_5 X_{15}$ 、 $X_6 X_{14}$

2.3. 研究方法

2.3.1. 逐步回归方法

设样本数据集为 (x_i, y_i) ， $i=1,2,\cdots,n$ ， n 为样本量。为了分析制丝过程烘丝段各参数对出料水分的影响关系，以表 1、表 2 中的因变量和自变量构建多元线性回归模型，表达式如下：

$$y_i = \alpha_i + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i \tag{1}$$

其中， β_j 为模型回归系数；随机误差向量 ε_i 满足 $E(\varepsilon_i) = 0$ ， $Var(\varepsilon_i) = \sigma^2$ 。

逐步回归方法的核心旨在处理包含大量自变量的数据集，通过精心筛选自变量子集，以构建既简洁高效又能对样本数据实现良好拟合效果的统计模型[12]。该方法采用迭代的方式引入和剔除变量，直至依据既定标准无法进一步优化模型。在逐步回归的过程中，存在三种主要策略：“向前”、“向后”以及“双向”。其中，“向前逐步回归”始于一个仅包含截距项的初始模型，随后逐一添加自变量，且每次添加均基于自变量对模型改进的贡献度进行评估；与之相反，“向后逐步回归”则从一个包含所有自变量的完整模型开始，逐步剔除对模型贡献较小的自变量；而“双向逐步回归”则融合了前两者的优势，灵活地在增加和减少变量的过程中寻求最优解。本文采用 R 语言中的 `step()` 函数，默认执行“双向”逐步回归，依据赤池信息准则自动选择最优模型[12]。

2.3.2. Lasso 方法

Lasso 方法的基本思想是在回归系数的绝对值之和小于一个常数的约束条件之下，使得残差平方和最小，即在约束条件 $\sum_{j=1}^p |\beta_j| \leq s$ 下，使得参数估计量 $(\hat{\alpha}, \hat{\beta})$ 满足如下：

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha, \beta)} \sum_{i=1}^n \left(y_i - \alpha_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (2)$$

Lasso 方法的惩罚函数形式为 $P_\lambda(|\beta_j|) = \lambda |\beta_j|$ ，其中 λ 为正则化参数。尽管引入 L1 正则化可能导致参数估计产生一定的偏差，但这一方法能够显著降低模型的预测方差，从而在整体上提高预测精度[13]。对于 Lasso 的求解问题，Efront 等在 2002 年提出了 lars (Least Angle Regression)算法[14]。在 Lasso 方法的模型估计过程中，R 语言中的 lars 程序包提供了两种主要的正则化参数选择方法：k 折交叉验证和 Cp (Complexity Parameter)统计量，其中，k 折交叉验证通过将数据集划分为 k 个子集，依次将每个子集作为验证集，其余子集作为训练集，从而评估模型的泛化性能；Cp 统计量则基于模型复杂度与拟合优度的权衡来选择最优模型。考虑到交叉验证方法在模型选择中的稳健性，本文采用 10 折交叉验证方法确定最优的正则化参数 λ ，并基于此获得最终选择的变量及其系数估计值。

2.3.3. 适应性 Lasso 方法

针对 Lasso 方法存在估计有偏的问题，Zou [11]在 2006 年提出了添加适应性 Lasso 惩罚项的方法，即在约束条件 $\sum_{j=1}^p w_j |\beta_j| \leq s$ 下，使得参数估计量 $(\hat{\alpha}, \hat{\beta})$ 满足式(2)。

适应性 Lasso 方法的惩罚函数形式为 $P_\lambda(|\beta_j|) = \lambda w_j |\beta_j|$ 。其中 $w_j = 1/(\hat{\beta}_j)^\gamma$ ，权重 w_j 通常基于 $\lambda \geq 0$ 时的岭回归估计值来计算，而 γ 则作为一个调整参数，用于控制惩罚的强度。

R 语言中的 msgps 程序包通过广义路径搜索算法实现了适应性 Lasso 方法，并提供了多种模型选择准则以确定最优正则化参数，包括 Mallows' Cp 统计量、偏差校正的 AIC (Corrected Akaike Information Criterion, AICc)、广义交叉验证准则(Generalized Cross Validation, GCV)以及贝叶斯信息准则(Bayesian Information Criterion, BIC)。其中，Mallows' Cp 统计量适用于模型复杂度与预测精度的权衡，AICc 在小样本情况下具有更好的校正效果，GCV 准则通过近似交叉验证来评估模型性能，而 BIC 准则在模型选择中倾向于选择更简洁的模型。根据现有文献研究[15]和实践经验表明，GCV 准则和 BIC 准则在变量选择中表现出更好的稳定性和可靠性，特别是在处理高维数据时。因此，本文采用 BIC 准则进行变量选择及其系数估计值的确定。

2.3.4. SCAD 方法

针对 Lasso 方法对变量系数过度压缩的问题，Fan 和 Li 在 2011 年进一步提出了添加 SCAD 惩罚项的方法，即在约束条件 $\sum_{j=1}^p |\beta_j| \leq s$ 下，使得参数估计量 $(\hat{\alpha}, \hat{\beta})$ 满足式(2)。SCAD 方法的惩罚函数形式为：

$$P_\lambda(|\beta_j|) = \begin{cases} \lambda |\beta_j| & \text{当 } \beta_j < \lambda \\ -\frac{|\beta_j|^2 - 2a\lambda|\beta_j| + \lambda^2}{2(a-1)} & \text{当 } \lambda \leq \beta_j \leq a\lambda \\ \frac{(a+1)\lambda^2}{2} & \text{当 } \beta_j > a\lambda \end{cases}$$

这里 $a > 2, \lambda > 0$ ，通常 a 取 3.7。本文通过 R 语言中的 ncvmreg 程序包中的 10 折交叉验证方法确定 SCAD 惩罚函数正则化参数 λ 的最优值。

2.4. 评价准则

2.4.1. AIC (Akaike Information Criterion)准则

赤池信息准则(AIC 准则)，由日本著名统计学家赤池弘次所创立，是一项广泛应用于模型选择与评估

的统计工具，尤其适用于那些可通过极大似然法进行参数估计的统计模型[16]。AIC 准则的统计量为：

$$AIC = n \ln(SSE_p/n) + 2p$$

其中， $SSE_p = \sum_{i=1}^n (y_i - y_{pi})^2$ ， y_{pi} 表示第 i 个观测值 y_i 基于模型预测得到的估计值。AIC 值越小表明模型拟合优度越高。

2.4.2. BIC (Bayesian Information Criterion) 准则

贝叶斯信息准则(BIC 准则)，与赤池信息准则(AIC 准则)在模型选择与评估方面扮演着类似的角色，但两者在处理模型复杂度与数据拟合度之间的权衡时有所不同。BIC 准则同样起源于统计学领域，旨在通过综合考虑模型的拟合优度与复杂度来避免过拟合现象，特别是在样本数量庞大的情况下，BIC 准则展现出了其独特的优势，能够有效地防止因模型精度过高而导致的复杂度过度增加[15]。BIC 准则的统计量为：

$$BIC = n \ln(SSE_p/n) + p \ln(n)/n$$

其中， $SSE_p = \sum_{i=1}^n (y_i - y_{pi})^2$ ， y_{pi} 表示第 i 个观测值 y_i 基于模型预测得到的估计值。BIC 值越小表明模型复杂度控制效果越好。

2.4.3. MSE (Mean Squared Error) 准则

均方误差准则(MSE 准则)作为一种简单而有效的模型评估工具，在统计学、机器学习、数据分析等领域得到了广泛的应用，为模型选择与优化提供有力的支持。MSE 准则通过对模型误差的平方和 SSE_p 进行平均化处理(基于其自由度 $n - p - 1$)，直观反映模型对数据集的整体拟合优劣程度[17]。MSE 准则的统计量为：

$$MSE = SSE_p / (n - p - 1)$$

其中， $SSE_p = \sum_{i=1}^n (y_i - y_{pi})^2$ ， y_{pi} 表示第 i 个观测值 y_i 基于模型预测得到的估计值。MSE 值越小表明模型预测精度越高。

3. 结果与分析

3.1. 不同变量选择方法的模型拟合结果分析

对样本数据进行标准化处理后，分别采用逐步回归、Lasso、适应性 Lasso 和 SCAD 等方法影响烘丝段出料水分的关键参数进行变量选择与建模分析，模型拟合结果如表 3 所示。由表 3 可知，不同方法对变量系数的估计值存在差异，但其符号方向基本一致。其中，变量 X_5 (热交换器薄膜阀阀门开度)和 X_{15} (热风温度)在逐步回归方法下的系数估计值为 0，而在 Lasso、适应性 Lasso 和 SCAD 方法下均被保留，表明逐步回归方法可能遗漏了这两个对出料水分具有实际影响的变量。此外，变量 X_{13} (I区筒壁温度)在逐步回归方法下系数显著不为 0，但在 Lasso 族方法下均被压缩至 0。进一步分析发现，变量 X_{13} 与变量 X_6 、 X_{14} 之间存在高度相关性(相关系数 > 0.9)，这验证了 Lasso 族方法在处理多重共线性问题上的显著优势。总体而言，与传统逐步回归方法相比，Lasso 族方法在变量选择结果上表现出明显差异，这不仅体现在所选变量的数量上，也反映在变量选择的有效性和稳定性方面。

当变量的系数估计值绝对值越大，表明该变量对出料水分的影响强度越高，其在水分控制中的重要性也越显著。根据表 3 的系数估计值对变量重要性进行排序，结果如表 4 所示。由表 4 可知，不同变量

Table 3. Model fitting results of various variable selection methods
表 3. 各变量选择方法的模型拟合结果

变量	参数	逐步回归	Lasso	适应性 Lasso	SCAD
X ₁	SX 蒸汽阀门开度	−0.4421***	−0.4475	−0.4518	−0.4685
X ₂	物料累计量	0.0764***	0.0740	0.0670	0.0811
X ₃	排潮阀门开度	0.1398***	0.0889	0.0986	0.1059
X ₄	循环风阀门开度	−0.1353	0.3244	0.3767	0.4315
X ₅	热交换器薄膜阀阀门开度	0.0000	0.5052	0.5138	0.5320
X ₆	筒壁二区蒸汽阀门开度	1.0355***	0.7519	0.8556	0.8897
X ₇	切叶丝含水率	0.3628***	0.4174	0.4208	0.4045
X ₈	工艺流量	−0.0487**	−0.0275	−0.0081	−0.0304
X ₉	膨胀单元蒸汽体积流量.	0.1445***	0.1222	0.1148	0.1300
X ₁₀	膨胀单元蒸汽质量流量.	−0.0452**	−0.0568	−0.0475	−0.0651
X ₁₁	负压	−0.1185***	−0.0725	−0.0902	−0.0954
X ₁₂	工艺气速度	0.3745***	0.2678	0.2904	0.3369
X ₁₃	I 区筒壁温度	0.4652***	0.0000	0.0000	0.0000
X ₁₄	II 区筒壁温度	−1.8299***	−1.2336	−1.3515	−1.3933
X ₁₅	热风温度	0.0000	0.0592	0.0525	0.0707
X ₁ X ₉	交互效应 1	0.0000	−0.0093	0.0000	−0.0025
X ₁ X ₁₀	交互效应 2	0.0000	0.0000	0.0000	0.0000
X ₃ X ₁₁	交互效应 3	0.0000	−0.0287	0.0000	−0.0073
X ₄ X ₁₁	交互效应 4	0.0000	0.0402	0.0000	0.0000
X ₅ X ₈	交互效应 5	0.0000	−0.0091	0.0000	0.0000
X ₅ X ₁₅	交互效应 6	0.0000	0.0046	0.0000	0.0000
X ₆ X ₁₄	交互效应 7	0.1169***	0.1070	0.1051	0.1125
X ₁ X ₉ X ₁₀	交互效应 8	0.0000	0.0005	0.0000	0.0000
X ₃ X ₄ X ₁₁	交互效应 9	0.0000	0.0591	0.0000	0.0043
X ₉ X ₁₀	交互效应 10	0.0000	0.0036	0.0000	0.0000
X ₃ X ₄	交互效应 11	−0.4961***	−0.2140	−0.1966	−0.2072

注：显著性水平：‘***’ 0.001；‘**’ 0.01；‘*’ 0.05；‘.’ 0.1；‘ ’ 1。

Table 4. Results of parameter screening and importance ranking
表 4. 变量选择及重要性排序结果

逐步回归方法	Lasso 方法	适应性 Lasso 方法	SCAD 方法
II 区筒壁温度	工艺气速度	II 区筒壁温度	II 区筒壁温度
交互效应 11	切叶丝含水率	切叶丝含水率	切叶丝含水率
I区筒壁温度	交互效应 7	工艺气速度	工艺气速度
工艺气速度	II 区筒壁温度	交互效应 11	交互效应 11

续表

切叶丝含水率	负压	膨胀单元蒸汽体积流量	膨胀单元蒸汽体积流量
膨胀单元蒸汽体积流量	膨胀单元蒸汽体积流量	交互效应 7	交互效应 7
负压	交互效应 3	负压	负压
交互效应 7	膨胀单元蒸汽质量流量	热风温度	热风温度
工艺流量	交互效应 11	膨胀单元蒸汽质量流量	膨胀单元蒸汽质量流量
膨胀单元蒸汽质量流量	热风温度	工艺流量	工艺流量
(热风温度)	交互效应 1	(I区筒壁温度)	交互效应 3
--	工艺流量	--	交互效应 9
--	交互效应 6	--	交互效应 1
--	交互效应 9	--	(I区筒壁温度)
--	交互效应 10	--	--
--	交互效应 5	--	--
--	交互效应 4	--	--
--	(交互效应 8)	--	--
--	(I区筒壁温度)	--	--
--	(交互效应 2)	--	--

注：排序从上至下表示重要性递减，括号表示被剔除的变量。

选择方法的结果在显著差异，但部分变量在四种方法中均被识别为重要变量。其中，II 区筒壁温度、切叶丝含水率、工艺气速度以及膨胀单元蒸汽体积流量和膨胀单元蒸汽质量流量在四种变量选择方法中均被识别为关键变量，表明这些参数对烘丝段出料水分具有显著影响，是水分控制的核心变量。交互效应 11 和交互效应 7 在多种方法中被一致选择，表明工艺参数间的交互作用对出料水分的影响不可忽视。当然，逐步回归方法选择的变量数量最少，但可能存在遗漏重要变量或交互效应的问题；Lasso 方法在变量选择上最为全面，能够识别更多的交互效应，但其选择的变量数量较多，可能导致模型复杂度较高；适应性 Lasso 方法和 SCAD 方法在变量选择上较为接近，能够平衡模型的简洁性与预测性能。

3.2. 多准则视角下的模型性能评估与比较

依据 AIC、BIC 和 MSE 三个评价指标分别对逐步回归、Lasso、适应性 Lasso 和 SCAD 等方法的模型性能进行评估，结果如表 5 所示。由表 5 可知，Lasso、适应性 Lasso 和 SCAD 等方法的 AIC 值显著低于逐步回归方法，其中 SCAD 方法的 AIC 值最小(-1174.735)，表明其模型拟合优度最高；Lasso 方法的

Table 5. Evaluation results of model performance
表 5. 模型性能评估结果

评价指标	逐步回归方法	Lasso 方法	适应性 Lasso 方法	SCAD 方法
AIC	-872.5410	-1133.2250	-1174.7120	-1174.7350
BIC	-922.4570	-1187.1330	-1128.6210	-1128.6440
MSE	0.6756	0.6023	0.5915	0.5914

BIC 值最低(-1187.133),表明其在模型复杂度和拟合优度之间取得了最佳平衡。适应性 Lasso 方法和 SCAD 方法的 BIC 值接近,但略高于 Lasso 方法,可能与变量选择的数量有关。Lasso、适应性 Lasso 和 SCAD 等方法的 MSE 值显著低于逐步回归方法,其中 SCAD 方法的 MSE 值最小(0.5914),表明其预测精度最高。因此,Lasso 族方法在模型拟合优度、预测精度和复杂度控制方面均显著优于传统逐步回归法,其中 SCAD 方法综合性能最优,适应性 Lasso 方法次之。

3.3. SCAD 模型在制丝水分控制中的应用

根据表 4 的 SCAD 模型变量选择及重要性排序结果,确定影响烘丝段水分控制的重要变量为:II 区筒壁温度、切叶丝含水率、工艺气速度、膨胀单元蒸汽体积流量、负压、热风温度、工艺流量等主效应参数,以及交互效应 7、交互效应 3、交互效应 9 和交互效应 1 等交互效应参数。其中,交互效应 11(排潮阀门开度与循环风阀门开度的交互效应)属于设备参数间的交互作用,在水分控制的工艺参数筛选中予以剔除。其余 4 个交互效应均涉及工艺参数,需将其对出料水分的影响转化为相应工艺参数独立影响的量化值。因此,烘丝段水分控制的重要工艺参数为:II 区筒壁温度、切叶丝含水率、工艺气速度、膨胀单元蒸汽体积流量、负压、热风温度、膨胀单元蒸汽质量流量和工艺流量。

3.3.1. II 区筒壁温度的影响分析

交互效应 7 涉及 II 区筒壁温度与筒壁二区蒸汽阀门开度两个参数的相互作用,以偏相关系数量化 II 区筒壁温度对出料水分的影响程度。根据表 3 的变量系统估计值,构建出料水分对 II 区筒壁温度的导数方程,如下:

$$\frac{\partial \text{出料水分}}{\partial \text{II 区筒壁温度}} = -1.3933 + 0.1125 \times \text{筒壁二区蒸汽阀门开度}$$

标准化后的筒壁二区蒸汽阀门开度取值区间为[-4.8450, 2.8140]。基于上述导数方程,计算得出 II 区筒壁温度对出料水分的偏相关系数范围为[-1.9384, -1.0767],表明当其它工艺参数固定为中心值时,出料水分与 II 区筒壁温度呈显著负相关关系;若 II 区筒壁温度每增加 1%,出料水分将减少约 1.08%至 1.94%。

3.3.2. 负压的影响分析

交互效应 3 和交互效应 9 涉及负压、排潮阀门开度和循环风阀门开度三个参数的相互作用,以偏相关系数量化负压对出料水分的影响程度。根据表 3 的变量系统估计值,构建出料水分对负压的导数方程,如下:

$$\frac{\partial \text{出料水分}}{\partial \text{负压}} = -0.0954 - 0.0073 \times \text{排潮阀门开度} + 0.0043 \times \text{排潮阀门开度} \times \text{循环风阀门开度}$$

标准化后的排潮阀门开度取值区间为[-1.5800, 4.4920],循环风阀门开度取值区间为[-8.2500, 1.3340]。基于上述导数方程,计算得出负压对出料水分的偏相关系数范围为[-0.1540, -0.0278]。表明当其它工艺参数固定为中心值时,出料水分与负压呈显著负相关关系;若负压每增加 1%,出料水分将减少约 0.03%至 0.15%。

3.3.3. 膨胀单元蒸汽体积流量的影响分析

交互效应 1 涉及膨胀单元蒸汽体积流量与 SX 蒸汽阀门开度两个参数的相互作用,以偏相关系数量化膨胀单元蒸汽体积流量对出料水分的影响程度。根据表 3 的变量系统估计值,构建出料水分对膨胀单元蒸汽体积流量的导数方程,如下:

$$\frac{\partial \text{出料水分}}{\partial \text{膨胀单元蒸汽体积流量}} = 0.13 - 0.4685 \times \text{SX 蒸汽阀门开度}$$

当标准化后的 SX 蒸汽阀门开度取值区间为[-4.7770, 3.8680]。基于上述导数方程，计算得出膨胀单元蒸汽体积流量对出料水分的偏相关系数范围为[-1.6822,2.3680]，表明若膨胀单元蒸汽体积流量每增加 1%，出料水分将反向减少约 0 至 1.68%或正向增加约 0 至 2.37%。

3.3.4. 其它关键工艺参数的影响分析

因其它五个工艺参数不涉及交互效应，其对出料水分的影响程度以变量系数估计值量化。其中，切叶丝含水率的系数估计值为 0.4045，然而切叶丝含水率作为烘丝段的外生变量，其值由上一道工序的加工生产决定，在烘丝段无法直接调节；工艺气速度的系数估计值为 0.3369，表明当其它工艺参数固定为中心值时，工艺气速度每增加 1%，出料水分将同向增加 0.34%；热风温度的系数估计值为 0.0707，表明当其它工艺参数固定为中心值时，热风温度每增加 1%，出料水分将同向增加 0.07%；膨胀单元蒸汽质量流量的系数估计值为-0.065，表明当其它工艺参数固定为中心值时，膨胀单元蒸汽质量流量每增加 1%，出料水分将反向减少 0.0651%；工艺流量的系数估计值为-0.0304，表明当其它工艺参数固定为中心值时，工艺流量每增加 1%，出料水分将反向减少 0.03%。

3.3.5. 关键工艺参数筛选及控制策略

根据表 6 所示的工艺参数影响程度排序结果，各工艺参数对烘丝段出料水分的影响程度存在显著差异。其中，II 区筒壁温度和膨胀单元蒸汽体积流量对出料水分的影响最为显著，切叶丝含水率和工艺气速度次之，而负压、热风温度、膨胀单元蒸汽质流量和工艺流量的影响相对较小。以 0.1 作为影响程度的分界点，筛选得到烘丝段的关键工艺参数依次为：II 区筒壁温度、膨胀单元蒸汽体积流量、切叶丝含水率和工艺气速度，其中前两者的影响尤为突出。

筒壁温度是影响烘丝段出料水分的核心参数之一。其作用机制主要体现在热传导效率的调控上。当筒壁温度升高时，热量通过筒壁向物料传递的速率加快，导致物料温度上升，从而加速物料内部水分的蒸发；反之，筒壁温度降低会减缓热传导过程，抑制水分蒸发。因此，II 区筒壁温度直接决定了物料的水分迁移速率和脱水效率。蒸汽体积流量通过调控膨胀单元内的热传递效率影响出料水分。当蒸汽流量增加时，单位时间内向物料传递的热量增多，物料温度快速上升，加速了内部水分的蒸发速率，从而降低出料水分含量；反之，蒸汽流量减少会削弱热传递效果，抑制水分蒸发，导致出料水分升高。当观测到出料水分偏低时，可优先考虑调控 II 区筒壁温度，适当降低筒壁温度以抑制物料内部水分的蒸发；若筒壁温度调整空间有限，可配合调整膨胀单元蒸汽体积流量，适当减少蒸汽流量以抑制水分蒸发，从而

Table 6. Ranking results of the importance of key process parameters
表 6. 关键工艺参数重要性排序结果

工艺参数	影响程度
II 区筒壁温度	[1.0767, 1.9384]
膨胀单元蒸汽体积流量	负向：[0, 1.6822]；正向：[0, 2.3680]
切叶丝含水率	0.4045
工艺气速度	0.3369
负压	[0.0278, 0.1540]
热风温度	0.0707
膨胀单元蒸汽质量流量	0.0651
工艺流量	0.0304

提高出料水分。这种多参数协同调控策略有助于实现出料水分的精准控制，确保烘丝段工艺的稳定性 and 产品质量的一致性。

4. 结论

① 通过四种变量选择方法的综合分析，II 区筒壁温度、切叶丝含水率、工艺气速度以及膨胀单元蒸汽体积流量和膨胀单元蒸汽质量流量均被识别为关键变量，表明这些参数对烘丝段出料水分具有显著影响，是水分控制的核心参数；② 在模型拟合优度、预测精度和复杂度控制方面，Lasso 族方法显著优于传统逐步回归方法，其中 SCAD 方法的综合性能表现最优；③ 基于 SCAD 方法筛选结果，烘丝段关键工艺参数依次为 II 区筒壁温度、膨胀单元蒸汽体积流量、切叶丝含水率和工艺气速度。

参考文献

- [1] 李艳, 卢游, 范翔, 等. 基于半参数回归模型的残差控制图[J]. 数理统计与管理, 2016, 35(4): 700-706.
- [2] 钟文焱, 陈晓杜, 马庆文, 等. 基于多因素分析的烘丝机入口含水率预测模型的建立与应用[J]. 烟草科技, 2015, 48(5): 67-73.
- [3] 谭超, 吴同, 覃鑫. 偏最小二乘组合后向区间选择在近红外定量建模中的应用[J]. 计算机与应用化学, 2008, 25(4): 509-512.
- [4] 卫盼盼, 吴祚友, 安银立, 等. 烟叶物理特性与打叶风分工艺参数的关系[J]. 烟草科技, 2014(8): 5-9.
- [5] 王明锋, 焦玉磊, 赵建华, 等. 部分烟气成分对卷烟感官舒适性的影响[J]. 中国烟草学报, 2014, 20(5): 12-18.
- [6] 李晋明, 端凯, 汪海, 等. 卷烟香型关键致香物筛选与香型定量判定模式[J]. 烟草科技, 2015, 48(12): 66-71.
- [7] Breiman, L. (1996) Heuristics of Instability and Stabilization in Model Selection. *The Annals of Statistics*, **24**, 2350-2383. <https://doi.org/10.1214/aos/1032181158>
- [8] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [9] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [10] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least Angle Regression. *The Annals of Statistics*, **32**, 407-409. <https://doi.org/10.1214/009053604000000067>
- [11] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [12] 吴喜之. 应用回归及分类——基于 R [M]. 北京: 中国人民大学出版社, 2016.
- [13] 谢佳春, 李兴绪. 明瑟收入函数模型变量选择方法比较研究[J]. 发展战略, 2016(1): 68-74.
- [14] 梁斌, 陈敏, 缪柏其, 等. 基于 LARS-Lasso 的指数跟踪及其在股指期货套利策略中的应用[J]. 数理统计与管理, 2011, 30(6): 1104-1113.
- [15] 吴喜之. 复杂数据统计方法——基于 R 的应用[M]. 北京: 中国人民大学出版社, 2012.
- [16] Daudin, J.J. (1986) Selection of Variables in Mixed-Variable Discriminant Analysis. *Biometrics*, **42**, 473-481. <https://doi.org/10.2307/2531198>
- [17] 刘立祥. 线性回归中自变量的选择与逐步回归方法[J]. 统计与决策, 2015(21): 80-82.