

基于自适应PCA算法对碳酸钙D3颗粒混合监测的研究

温 婷, 周末千, 王 春, 游冬梅, 陈 东

华润三九医药股份有限公司, 广东 深圳

收稿日期: 2025年8月28日; 录用日期: 2025年9月15日; 发布日期: 2025年11月7日

摘 要

目的: 利用近红外光谱技术结合自适应主成分分析(Principal Components Analysis, PCA)算法实现批次差异较大碳酸钙D3颗粒混合终点判断。方法: 利用滚动PCA模块空间变化计算混合过程中模块之间的马氏距离, 反映碳酸钙D3颗粒混合过程变化趋势, 并设定阈值指导混合终点。结果: 通过自适应PCA空间结构进行降维并预测当前混合点位与中心点位的马氏距离, 通过结合混合颗粒情况确定30个光谱作为1个块, 并同时跟进2组马氏距离趋势, 明确控制阈值为1.73, 为验证方法可靠性, 采用HPLC方法对维生素D3活性成分进行含量检测, 实测均匀度结果与自适应PCA算法判定结果一致。结论: 使用自适应PCA算法可用于碳酸钙D3颗粒混合过程监控并进行终点判断, 本方法可随颗粒体系的变化灵活调整参数, 从而快速调控混合终点, 以期颗粒中间体或终产品的混合均匀度监测提供新思路。

关键词

自适应算法, 混合工艺, 碳酸钙D3颗粒, 均匀度判断

Research on Monitoring the Mixing of Calcium Carbonate D3 Granules Based on Adaptive PCA Algorithm

Ting Wen, Moqian Zhou, Chun Wang, Dongmei You, Dong Chen

China Resources Sanjiu Medical & Pharmaceutical Co., Ltd., Shenzhen Guangdong

Received: August 28, 2025; accepted: September 15, 2025; published: November 7, 2025

Abstract

Objective: To utilize near-infrared spectroscopy technology combined with the adaptive principal

文章引用: 温婷, 周末千, 王春, 游冬梅, 陈东. 基于自适应 PCA 算法对碳酸钙 D3 颗粒混合监测的研究[J]. 药物化学, 2025, 13(4): 305-314. DOI: 10.12677/hjmce.2025.134031

components analysis (PCA) algorithm to determine the mixing endpoint of calcium carbonate D3 granules with significant batch differences. **Methods:** The space variation of the rolling PCA module was used to calculate the Mahalanobis distance between modules during the mixing process, reflecting the changing trend of the mixing process of calcium carbonate D3 granules, and a threshold was set to guide the mixing endpoint. **Results:** The dimensionality reduction was carried out through the adaptive PCA spatial structure to predict the Mahalanobis distance between the current mixing point and the center point. By combining the conditions of mixing particles, 30 spectra were selected as one block, and the two sets of Mahalanobis distance trends were simultaneously followed up. The control threshold was clearly set at 1.73. To verify the reliability of the method, the content of vitamin D3 was detected by HPLC. The measured uniformity results showed consistency with the determination results of the adaptive PCA algorithm. **Conclusion:** The adaptive PCA algorithm can be used to monitor the mixing process of calcium carbonate D3 granules and determine the mixing endpoint. This method can flexibly adjust parameters according to the changes of the particle system to quickly control the mixing endpoint, with the aim of providing a new idea for the monitoring of the mixing uniformity of particle intermediates or final products.

Keywords

Adaptive Algorithm, Mixing Process, Calcium Carbonate D3 Granules, Uniformity Judgment

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

在现代化的工业生产流程中,对于共混工艺这种对物料混合均匀度要求极高的环节,精准地判断工艺何时达到终点至关重要。传统的生产模式以分批进行、逐批检验放行为主,每一生产步骤完结后,必须先进行取样并送至实验室进行离线检测,合格后方可启动后续步骤,这种模式下的生产周期普遍较长,从几天到几周不等,显著拖慢了整体效率[1]-[3]。同时,离线检测的取样方式存在先天不足,无法完全捕捉整批产品的特性,容易造成产品稳定性和均一性不佳的问题,最终影响产品的安全性和质量可靠性[4][5]。

利用近红外(Near Infrared, NIR)光谱技术对药物生产过程进行终点判断有助于及时、准确地识别过程终点,降低生产成本,提高原辅料利用率,保证产品质量的均一稳定[6]-[8]。利用物质独特的“指纹”——光谱性质,通过实时监测光谱信号的变化,可以洞察混合物内部成分分布的动态演变[9][10]。目前,行业内普遍采用的常规技术是基于移动块分析(Moving Block Analysis)的光谱终点检测方法[11][12]。这种分析方法的核心思想是,将连续采集的光谱数据序列分割成一系列连续的、长度固定的“数据块”,然后对每个块内的数据进行统计处理,以捕捉信号的整体趋势和波动特性[13]。常用的统计指标包括(1)移动块标准偏差(Moving Block Standard Deviation, MBSD):衡量每个数据块内光谱信号的离散程度,反映混合物成分的均匀性[14];(2)移动块平均值(Moving Block Mean, MBM):追踪每个数据块光谱信号的中心位置,指示混合物整体性质的平均水平[15];(3)移动块相对标准偏差(Moving Block Relative Standard Deviation, MB-RSD):通过衡量块内数据的离散程度相对于其平均水平的比例大小,相较于 MBSD 标准化了波动程度,便于不同量级数据间的比较[16]。

然而,这些统计指标在实际运用过程中,其局限性也逐渐显现。MBSD、MBM 和 MB-RSD 是单一的统计量,往往难以全面、稳健地刻画复杂的混合过程。对数据中的异常值或过程本身的微小波动过于

敏感,导致终点判断不够稳定可靠,即缺乏“鲁棒性”[17][18]。通常需要大量的理化数据去验证一个合理的阈值,这就意味着是基于历史数据进行框定的条件,当批次之间的颗粒如水分、粒度分布等特性差异较大时,用其中几批的混合终点光谱作为标准来预测其他批次易存在模型偏差,此时建立的模型所需样本量大且极具不稳定性。

因此,本研究尝试通过建立自适应主成分分析(自适应 PCA)模型来在线监测碳酸钙 D3 颗粒的混合均匀度,其基本原理是通过在连续滚动的局部光谱区间窗口上反复应用 PCA,建立多个 PCA 模型,来估计和追踪数据协方差结构随时间的变化(批次之间独立运行算法),同时根据马氏距离的阈值限判断混合的终点。

2. 材料与方法

2.1. 物料与仪器

物料来源:碳酸钙颗粒自制(华润三九医药股份有限公司),维生素 D3 颗粒自制(华润三九医药股份有限公司,过 50 目药典筛)

近红外光谱仪: MicroNIR-PAT-U;

高剪切湿法制粒机: 奥星, 4L 制粒锅;

高效液相色谱仪(HPLC): 日本岛津, LC-20AT。

2.2. 方法

2.2.1. 近红外光谱采集

采用 MicroNIR-PAT-U, 空气做背景,采用漫反射方式采集混合过程中光谱数据,光谱 908.1-1676.2 nm,积分时间 5.1 ms,扫描次数 100 次,扫描间隔 2 s/1 次。

光谱处理使用 Unscrambler X 软件,自适应算法在 Python 软件平台上完成。

2.2.2. 自适应 PCA 算法运行

通过混合过程采集的光谱数据形成每批次的模型,运用自适应 PCA 算法用于动态过程监测与终点检测。其核心技术路线是通过滚动更新 PCA 模型,计算后续模型到前一块的距离度量,对动态过程的光谱数据进行实时分析,判断过程是否达到稳态或终点。

具体操作:

使用湿法制粒机进行碳酸钙 D3 颗粒的混合工艺,混合比例为碳酸钙颗粒:维生素 D3 颗粒 = 150:1,设置搅拌桨转速为 150 rpm,开始混合过程并采用 MicroNIR-PAT-U 探头收集光谱数据。将按时间序列的全过程光谱划分为连续的光谱块(Block),每个块包含固定数量的光谱(如设定块尺寸为 n ,则块 1 包含光谱 $S_1 \sim S_N$,则块 1.1 包含光谱 $S_{N+1} \sim S_{2N+1}$),采用滑动窗口方式生成后续块,如块 1 包含 $S_1 \sim S_N$,块 2 包含 $S_2 \sim S_{N+1}$ 等,实现数据的滚动处理。对不同移动块的光谱数据进行预处理及主成分分析,确定当前空间使用的预处理方法、特征向量、均值向量、标准差等,计算当前块组的后一个块的重构数据及距离度量(如计算块 1.1 到块 1 的马氏距离或霍特林 T^2 等),当距离度量达到设定阈值标准,则混合达到稳态或终点。阈值设定可根据单因素指标或多因素指标判定,如:同时满足设定要求连续 3 个块的马氏距离均 < 阈值,或霍特林 T^2 超过阈值的次数 < 块尺寸的 10% (如 3 次/30 光谱)等。若达到阈值或设定标准,则进行块 2 组的判定,直至达到混合稳态或终点(原理流程图见图 1)。

计算方法:

1. 建立基础块 PCA 模型

训练集光谱数据矩阵 $X \in R^{m \times p}$, n : 训练样本数; p : 光谱波长点数(变量维数)

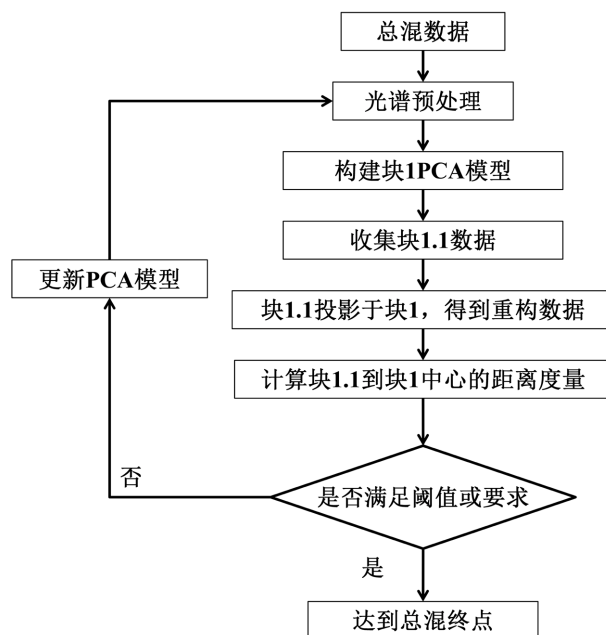


Figure 1. Flowchart of the adaptive PCA algorithm principle
图 1. 自适应 PCA 算法原理流程图

(1) 中心化: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i (\bar{x} \in R^p)$;

(2) 中心化数据: $X_c = X - 1_n \bar{x}^T$;

(3) 计算协方差矩阵: $C = \frac{1}{n-1} X_c^T X_c (C \in R^{p \times p})$;

(4) 特征分解: $C = P \Lambda P^T$ 。

$P \in R^{p \times p}$: 特征向量矩阵(列向量为单位特征向量)。

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$: 特征值对角矩阵($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$)。

(5) 选择主成分数 k :

$P_k = [p_1, p_2, \dots, p_k] \in R^{p \times k}$;

$\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$: 前 k 个特征值。

2. 重构块光谱数据

(1) 新光谱数据矩阵 $X_{new} \in R^{m \times p}$, m : 新样本数;

(2) 中心化: $X_{c,new} = X_{new} - 1_m \bar{x}^T$;

(3) 投影到主成分空间: $T_{new} = X_{c,new} P_k (T_{new} \in R^{m \times k})$

(4) 重构光谱数据: $\hat{X}_{c,new} = T_{new} P_k^T$; $\hat{X}_{new} = \hat{X}_{c,new} + 1_m \bar{x}^T (\hat{X}_{new} \in R^{m \times p})$ 。

3. 计算每个块的马氏距离的均值

$$MD_j = \sqrt{t_{new,j}^T (\Lambda_k^{-1}) t_{new,j}}$$

$t_{new,j}$: 第 j 个新样本的得分向量(T_{new} 的第 j 行)。

$\Lambda_k^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_k)$: 特征值逆矩阵

$$\overline{MD} = \frac{1}{m} \sum_{j=1}^m MD_j$$

注：符号含义见表 1。

Table 1. Symbol meanings
表 1. 符号含义

符号	含义	维度	符号	含义	维度
n	训练样本数	标量	p	光谱波长点数	标量
m	新样本数	标量	k	主成分数	标量
X	训练集光谱数据矩阵	$n \times p$	$\bar{x}$	训练集均值向量	$p \times 1$
X_c	中心化训练数据	$n \times p$	C	协方差矩阵	$p \times p$
P	全部特征向量矩阵	$p \times p$	Λ	全部特征值对角矩阵	$p \times p$
P_k	前 k 个载荷向量	$p \times k$	Λ_k	前 k 个特征值对角矩阵	$k \times k$
X_{new}	新批次光谱数据	$m \times p$	$X_{c,new}$	中心化新数据	$m \times p$
T_{new}	新数据得分矩阵	$m \times k$	$\hat{X}_{new}$	重构光谱数据	$m \times p$
$t_{new,j}$	第 j 个新样本的得分向量	$1 \times k$	MD_j	第 j 个样本的马氏距离	标量
$\overline{MD}$	马氏距离的均值	标量	/		

2.2.3. 自适应 PCA 算法验证

为验证建立的自适应 PCA 算法的准确性，当自适应 PCA 算法显示混合达到终点时，采用取样器从湿法制粒锅不同位置取样 10 个点(表层 5 个取样点，底层 5 个取样点；每个取样点 1 个正常样、1 个备用样)，使用 HPLC 法测定 10 个取样点中维生素 D3 含量，并计算 RSD 值，当 $RSD \leq 5\%$ 时，表明混合均匀度达到要求，印证模型方法判断终点的准确性。HPLC 检测条件如下表 2 所示：

Table 2. HPLC conditions for the determination of vitamin D3 content
表 2. 维生素 D3 含量测定 HPLC 色谱条件

色谱柱	Ultimate LP-C18, 4.6 * 250 mm, 5 μ m	流速	1.5 mL/min	检测波长	265 nm
进样体积	100 μ L	柱温	30 $^{\circ}$ C	运行时间	35 min
流动相	以水 - 乙腈(40:60)为流动相 A，以乙腈 - 甲醇(95:5)为流动相 B				
洗脱条件	时间(min)	A%		B%	
	0	100		0	
	5	0		100	
	25	0		100	
	26	100		0	
	35	100		0	

3. 结果与讨论

3.1. 光谱预处理

共收集 3 批，碳酸钙 D3 颗粒混合过程采集的近红外原始光谱如图 2(a)所示，预处理方法选用 SNV，以消除由于颗粒分布不均导致的基线漂移，预处理后的光谱如图 2(b)所示。

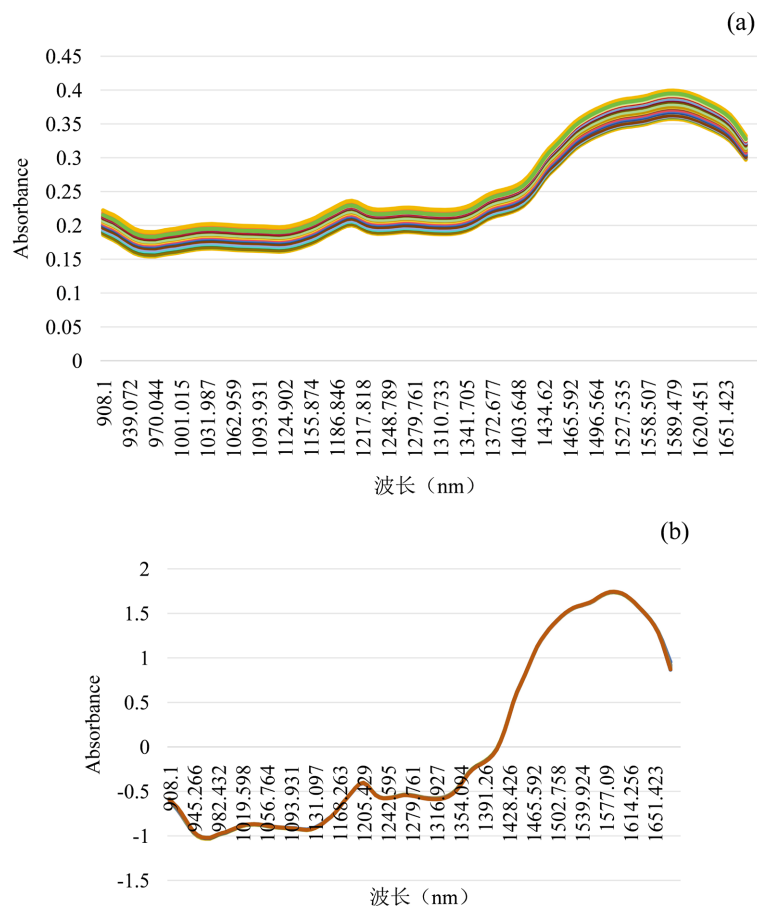


Figure 2. (a) Original NIR spectrogram, and (b) NIR spectrogram after SNV pre-processing

图 2. (a) 原始 NIR 光谱图; (b) 经 SNV 预处理后的 NIR 光谱图

3.2. 自适应 PCA 算法中参数的确定

采用自适应 PCA 算法判定碳酸钙 D3 混合终点时, 终点距离判断方法使用马氏距离计算, 研究需求确定阈值为 DM, 主成分数 k 和块大小 n 。工业生产中针对马氏距离的阈值 DM 大多设定为 3 (由于投影点有 99.7% 的概率由 PCA 描述, 即错判的概率是 0.3%), 针对不同的生产混合比例, 需结合当前生产实际情况对阈值进行调整, 本研究中碳酸钙空白颗粒: 维生素 D3 颗粒 = 150:1, 混合体系中维生素 D3 颗粒整体占比较小, 达到混合均匀终点时, 光谱仅在很小的范围内随机波动, 将 DM 设定为 3 无法精准监测混合终点, 因此将阈值实际设定为 1.73, 以识别混合比例较大的样品。主成分数的选择需根据降维累计贡献程度判定, 由于本试验的光谱数据质量相对较好, 前两个主成分即可解释数据 95% 以上的变异, 因此主成分数选择 2。在碳酸钙 D3 混合过程中, 每条光谱代表当前混合状态, 块的大小决定建模数据多少及块与块之间马氏距离大小, 为代表光谱的代表性和灵敏度, 本研究的块大小选择为 30 (在 PCA 建模分析中, 光谱数据决定模型稳定性, 块设定为 30 表示后一个光谱与前一分钟混合过程的变化), 即每次使用 30 条连续光谱进行分析, 并依次使用新的光谱代替最早的光谱进行迭代滚动分析。

为使碳酸钙 D3 混合终点的判断更为精确, 需考察连续两组数据间均达到混合终点标准, 具体参照 2.2.2 中的算法, 计算块 1 与块 1.1 之间距离(记为 D1), 计算块 1 与块 1.2 之间距离(记为 D2), 具体示意图如图 3。

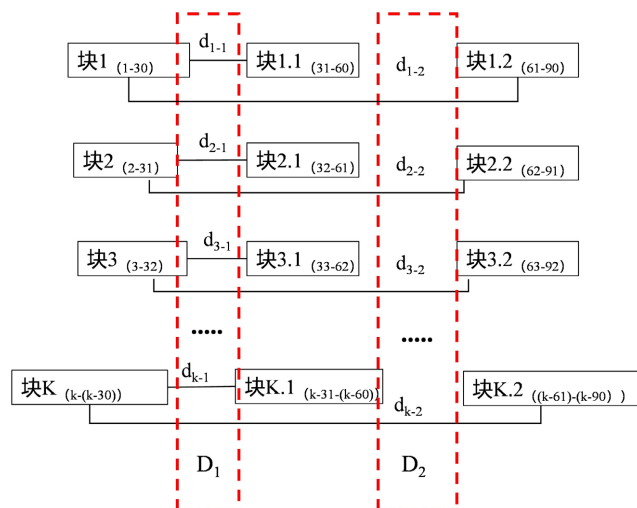


Figure 3. Schematic diagram of adaptive PCA algorithm

图 3. 自适应 PCA 算法示意图

3.3. 自适应算法终点判断结果分析

本研究对碳酸钙 D3 颗粒(批号: 20250730)的混合过程进行监控, 采用传统方法 MBSD 方法进行监控, 结果如图 4 所示, MBSD 由于混合颗粒差异较大, 无法准确跟进混合变化过程, 方法不适用。采用自适应 PCA 算法进行在线监测, 结果见图 5 所示, 在第 535 块时 D1 与 D2 均在阈值线下, 为确定生产阶段达到稳定状态, 当 D1 与 D2 均 <1.73 时, 计算 20 个块, 且 90%的块位于阈值线下, 即可认为达到碳酸钙 D3 颗粒混合终点时间。为验证数据分析方法准确性, 将于第 535 块(即混合开始第 15 min)时进行取样, 采用液相测定各取样点($n=10$)样品中维生素 D3 的含量(根据 2.2.3 项下检测条件), 并计算 RSD 值为 $1.47\% \leq 5\%$, 证明本算法可用于混合终点判断。

为考察方法稳健性, 使用与上述批次差异较大的两个批次(批号: 20250617、20250722)如图 6 所示进行验证。验证结果如图 7 所示, 按照 2.2.3 项下检测条件, 测定的维生素 D3 含量 RSD 值分别为 1.76%和 3.24%, 均低于 5%阈值, 结果表明, 自适应 PCA 算法适用于批次间差异较大的碳酸钙 D3 颗粒混合均匀度判断。

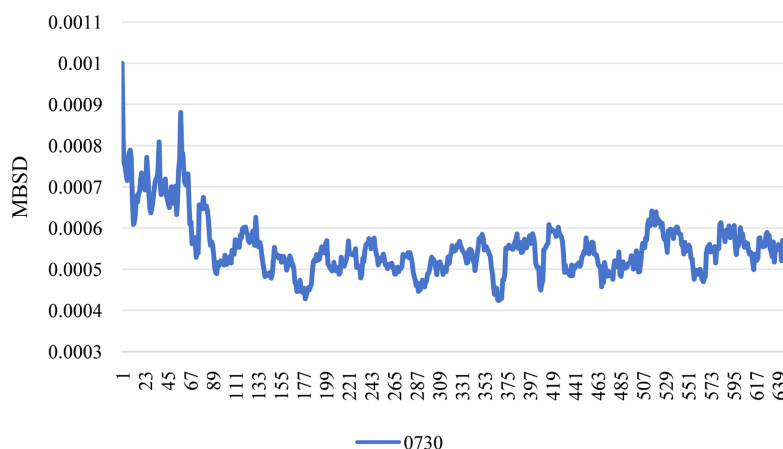


Figure 4. Trend chart of hybrid termination judgment results of MBSD algorithm

图 4. MBSD 算法混合终点判断结果趋势图

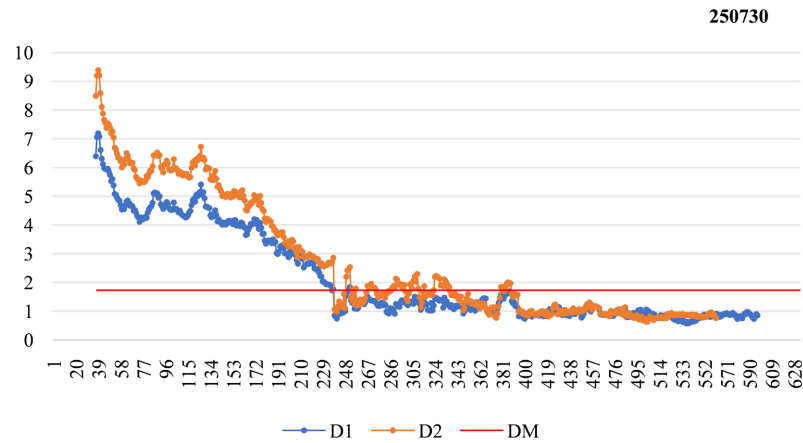


Figure 5. Trend chart of adaptive PCA hybrid endpoint judgment results
图 5. 自适应 PCA 混合终点判断结果趋势图

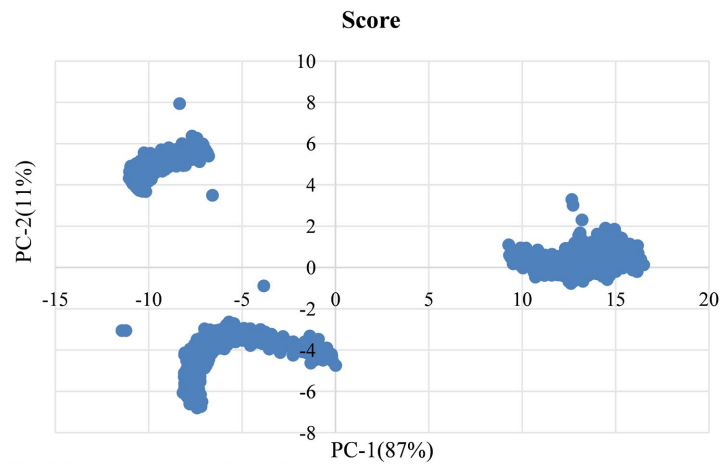
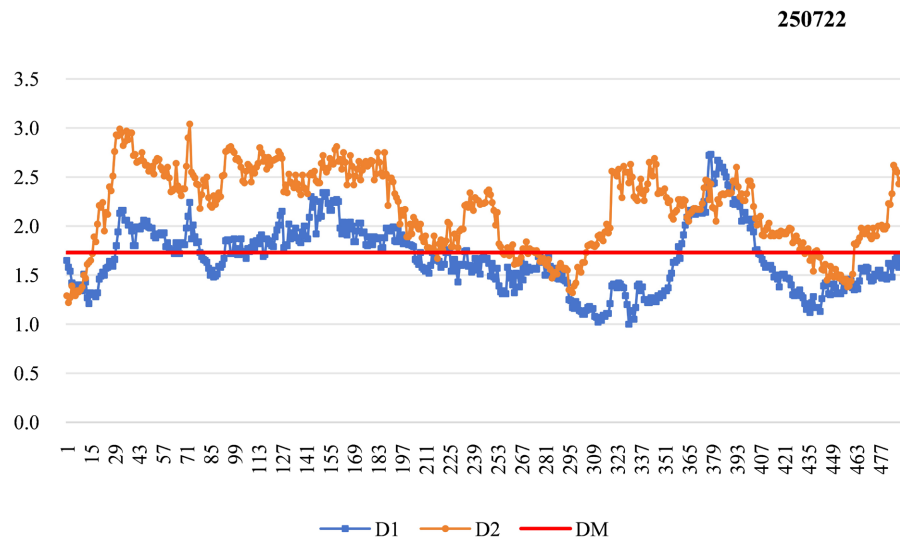


Figure 6. Comparison of PCA differences among three batches of calcium carbonate granules
图 6. 三批次碳酸钙颗粒 PCA 差异对比



250617

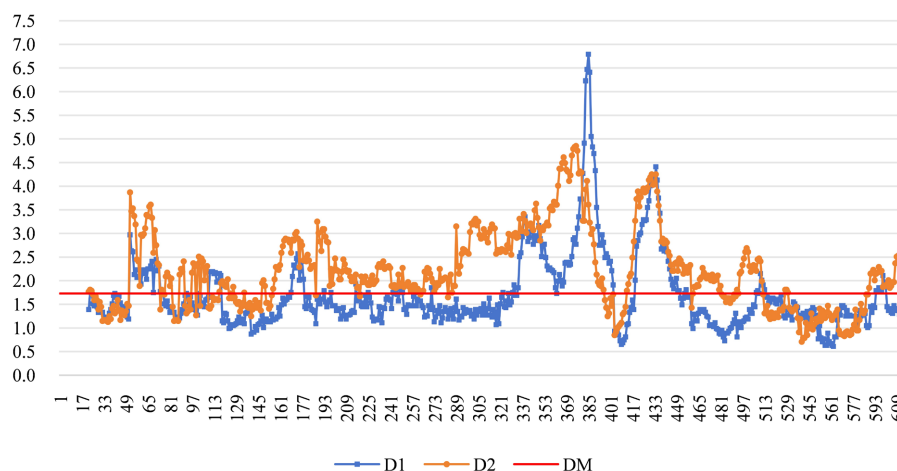


Figure 7. Verification group adaptive PCA mixed endpoint judgment results
图 7. 验证组自适应 PCA 混合终点判断结果图

4. 结论

本研究以碳酸钙 D3 混合过程分析为研究对象, 针对批次间差异较大样本混合过程终点的判断问题构建新方法, 使用液相实际测定三批产品混合终点时不同取样点的维生素 D3 含量并计算 RSD 值, 验证算法混合终点判断结果的准确性。本自适应 PCA 算法可用于混合过程均一性在线监测, 其终点判断仅依赖实时采集的光谱数据, 无需预校准或历史样本, 通过滚动更新模型, 适应过程参数的缓慢变化, 避免过早误判终点, 相较传统判别方法更加准确、全面, 可灵活快速判断混合终点, 确保产品均匀度符合质量要求。

参考文献

- [1] 杨真真, 蔡建峰, 黄秋萍, 等. HPLC 法应用于桃仁通痹丸生产工艺中混合均匀度的研究[J]. 海峡药学, 2024, 36(9): 28-32.
- [2] 朱飞, 辛洪亮. 不同处方工艺对非那雄胺片混合均匀度与含量均匀度的影响[J]. 广州化工, 2025, 53(5): 137-140, 152.
- [3] Florian-Algarin, M. and Méndez, R. (2015) Effect of Material Properties and Design Parameters on the Final Blend Uniformity Using Experimental and Simulation Results. *Powder Technology*, **276**, 156-165. <https://doi.org/10.1016/j.powtec.2015.02.024>
- [4] 吴思俊, 仇萍, 彭晓珊, 等. 在线监测技术在药物混合过程中的应用进展[J]. 中南药学, 2019, 17(9): 1425-1430.
- [5] Moes, J.J., Ruijken, M.M., Gout, E., Frijlink, H.W. and Ugwoke, M.I. (2008) Application of Process Analytical Technology in Tablet Process Development Using NIR Spectroscopy: Blend Uniformity, Content Uniformity and Coating Thickness Measurements. *International Journal of Pharmaceutics*, **357**, 108-118. <https://doi.org/10.1016/j.ijpharm.2008.01.062>
- [6] 付立娟, 邢成, 边雨, 等. 香砂胃苓丸混合终点在线监测及质量评价近红外模型的建立[J]. 安徽农业科学, 2024, 52(4): 153-156.
- [7] 田莹, 赵洁, 瞿海斌. 基于近红外光谱技术的银杏叶滴丸熔融物料混合过程监测方法的开发[J]. 中草药, 2023, 54(13): 4137-4143.
- [8] 吴思俊. 基于近红外光谱技术的中药制药工艺终点判断方法研究[D]: [硕士学位论文]. 天津: 天津中医药大学, 2021.
- [9] 叶世著, 吴艾璟. 近红外光谱在物料混合均匀度判定中的应用进展[J]. 广东化工, 2019, 46(15): 106-107, 112.
- [10] Dong, W., Wang, X., Huang, Z., Ye, C., Wang, T., Zhang, H., et al. (2025) On-line Monitoring of Guizhi Fuling Capsules

- and Tablets Dissolution Behavior Using Near-Infrared Spectroscopy Combined with Chemometrics. *Journal of Analysis and Testing*, **9**, 240-249. <https://doi.org/10.1007/s41664-025-00349-y>
- [11] 苏婷, 姜文月, 李亚东, 等. 声光可调-近红外光谱法快速判断精芪双参胶囊的混合终点及测定黄芪甲苷的含量[J]. 中国药房, 2018, 29(12): 1616-1620.
- [12] Storme-Paris, I., Clarot, I., Esposito, S., Chaumeil, J.C., Nicolas, A., Brion, F., *et al.* (2009) Near Infrared Spectroscopy Homogeneity Evaluation of Complex Powder Blends in a Small-Scale Pharmaceutical Preformulation Process, a Real-Life Application. *European Journal of Pharmaceutics and Biopharmaceutics*, **72**, 189-198. <https://doi.org/10.1016/j.ejpb.2008.11.002>
- [13] 林兆洲, 杨婵, 徐冰, 等. 中药混合过程终点在线判定方法研究[J]. 中国中药杂志, 2017, 42(6): 1089-1094.
- [14] 彭秋实. 近红外光谱技术在布洛芬复方片制备工艺中的应用研究[D]: [硕士学位论文]. 贵阳: 贵州大学, 2018.
- [15] 白玉莹, 王石, 张博涵, 等. MicroNIR 技术在配方奶粉干混均匀性上的应用[J/OL]. 现代食品科技: 1-12. <https://doi.org/10.13982/j.mfst.1673-9078.2025.7.0404>, 2025-07-30.
- [16] 李蓓蓓. 基于烟草近红外光谱特征的变量选择方法研究及应用[D]: [硕士学位论文]. 郑州: 中国烟草总公司郑州烟草研究院, 2025.
- [17] 童枫, 徐芳芳, 闫逸伦, 等. 热毒宁注射液金银花和青蒿(金青)萃取过程中固形物含量近红外光谱在线监测模型的建立及萃取终点判断研究[J]. 中草药, 2024, 55(19): 6555-6565.
- [18] 陈厚柳. 银杏叶提取和层析过程在线质量控制方法研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2015.