

· 论著 ·

## 硫酸羟氯喹颗粒水分近红外光谱在线定量模型的建立

柯 樱<sup>1</sup>, 朱振明<sup>2</sup>, 张烁阳<sup>3</sup>, 王薇青<sup>3</sup>, 陆 峰<sup>3</sup> (1. 上海医药集团股份有限公司, 上海 200020; 2. 上海上药中西制药有限公司, 上海 201806; 3. 海军军医大学, 上海 200433)

**【摘要】** 目的 为实时检测硫酸羟氯喹颗粒在流化床干燥过程中的水分含量变化, 建立颗粒水分的在线近红外光谱定量模型。方法 物料颗粒在流化床的干燥过程中, 实时取样并用水分测定仪测量颗粒水分, 采用多元散射校正(multiplicative signal correction, MSC)、一阶导数和 Karl Norris 平滑的光谱预处理方法, 选择近红外 4 935 ~ 5 336  $\text{cm}^{-1}$  和 6 911 ~ 7 297  $\text{cm}^{-1}$  两个波段, 运用偏最小二乘回归(partial least squares, PLS)建立基于近红外光谱的水分定量分析模型。结果 所建模型的校正误差均方根(RMSEC)为 0.408, 相关系数  $R_c$  为 0.952 9。预测误差均方根(RMSEP)为 0.435, 相关系数  $R_p$  为 0.936 6, 验证集标准偏差与预测标准偏差的比值(RPD)为 5.18。并将该模型应用于车间生产过程中,  $t$  检验结果表明, 预测值与参考值之间无显著性差异。结论 该法所建立的在线水分定量模型准确度较高且较为可靠稳定, 该模型可应用于生产过程, 以在线监测物料颗粒的水分变化。

**【关键词】** 近红外光谱分析技术; 硫酸羟氯喹颗粒; 水分含量; 在线定量模型; 干燥过程

**【中图分类号】** O657.33; R284.1 **【文献标志码】** A **【文章编号】** 1006-0111(2021)01-0023-06

**【DOI】** 10.12206/j.issn.1006-0111.202007128

## Establishment of online quantitative model for moisture content determination of hydroxychloroquine sulfate particles by near infrared spectroscopy

KE Ying<sup>1</sup>, ZHU Zhenming<sup>2</sup>, ZHANG Shuoyang<sup>3</sup>, WANG Weiqing<sup>3</sup>, LU Feng<sup>3</sup> (1. Shanghai Pharmaceuticals Holding Co., Ltd., Shanghai 200020, China; 2. Shanghai SPH Zhongxi Pharmaceutical Co., Ltd., Shanghai 201806, China; 3. Naval Medical University, Shanghai 200433, China)

**【Abstract】** **Objective** To establish an online quantitative analysis model for moisture content assay of hydroxychloroquine sulfate particles by near infrared (NIR) spectroscopy. **Methods** The NIR spectra were collected in real time when the material particles were dried in the fluidized bed. Meanwhile the water content of the particles was measured with the standard moisture tester. The multiplicative signal correction (MSC) and first derivative followed by Karl Norris smoothing were used for spectra pretreatment. Two spectral range (4 935–5 336  $\text{cm}^{-1}$  and 6 911–7 297  $\text{cm}^{-1}$ ) were selected for the quantitative model with the partial least squares (PLS) regression. **Results** The quantitative calibration model had good correlation coefficients with  $R_c$  value=0.952 9 and  $R_p$  value=0.936 6. The root mean square error of calibration (RMSEC) was 0.408 and the root mean square error of prediction error (RMSEP) was 0.435. The ratio of standard deviation of validation set to prediction standard deviation (RPD) was 5.18. There was no significant difference between the predicted value and the reference value by  $t$  test when the established model was applied in large-scale production. **Conclusion** The online model established for monitoring water content has high accuracy and stability, which can be applied in industrial scale process to monitor the particle moisture in real time.

**【Key words】** near infrared spectroscopy; hydroxychloroquine sulfate particle; water content; online quantitative model; drying process

过程分析技术 (process analytical technology, PAT) 通过对关键质量数据包括原始物料质量、

**【基金项目】** 国家重点研发计划: 口服固体制剂生产过程实时监测与控制关键技术、应用及相关监管法规研究(2017YFF0210100)

**【作者简介】** 柯 樱, 硕士, 正高级工程师, 研究方向: 药物化学, Email: key@sphchina.com, Tel: 13501748373

**【通信作者】** 陆 峰, 教授, 博士生导师, 研究方向: 药物、生物光谱分析研究, Email: fenglu@smmu.edu.cn

中间物料质量和工艺过程参数进行实时监测以确保成品质量<sup>[1]</sup>。其中, 近红外光谱 (near infrared spectroscopy, NIRS) 分析技术是 PAT 的核心技术之一, 其独特优势在于测试方便、分析速度快、分析效率高, 非常适用于在线分析<sup>[2]</sup>。目前该技术在海外药企已有很多应用, 如对流化过程中干燥终点的控制<sup>[3]</sup>、监测包衣过程<sup>[4]</sup>、测定片剂中原料药和辅料的含量<sup>[5]</sup>等。

硫酸羟氯喹是临床上治疗类风湿性关节炎、系统性红斑狼疮的常用药物<sup>[6-8]</sup>。硫酸羟氯喹制粒工艺为湿法制粒,而湿法制粒的产品质量控制主要在制粒过程中。为缩小批间差异和改善工艺,利用 PAT 技术建立定量或定性模型以检测颗粒水分、混合均匀度和粒径等一系列指标,而水分含量既影响粒径分布也影响混匀程度,故需要优先测量<sup>[9]</sup>。目前,硫酸羟氯喹颗粒在干燥过程中使用快速水分测定仪测定水分含量,以监测干燥终点。这种分析方法在干燥过程中取样测定水分,会带来人为干扰因素,且水分测定仪的测量时间较长,在等待测量结果时,流化床内的物料水分仍在不断变化。使用 NIR 分析技术对硫酸羟氯喹颗粒的水分含量进行实时监测,将有助于提高硫酸羟氯喹的生产效率、降低因等待测量结果时物料水分仍在不断降低带来的生产风险,以及提高产成品的质量,有助于向生产全自动化发展。目前,国内使用 NIR 分析技术监测物料水分含量的报道以离线方式居多<sup>[10-13]</sup>。课题组之前建立了硫酸羟氯喹颗粒离线水分定量模型<sup>[10]</sup>,本研究在此基础上先在小试部分建立在线水分定量模型,并在试运行情况良好的基础上将模型转移到车间生产线。结果表明该模型能满足生产过程在线监测硫酸羟氯喹颗粒水分的需求。

## 1 仪器与材料

Thermo Antaris MX 近红外光谱仪、TQ Analyst 9.5 软件(美国 Thermo Fisher 公司),HX204 卤素水分测定仪(瑞士 Mettler-Toledo 公司),硫酸羟氯喹颗粒(上海上药中西制药有限公司),MATLAB 2014a(美国 MathWorks 公司),DPL-0.5 型多功能

制粒/包衣机(重庆精工制药机械有限责任公司),DG400 流化床(上海东富龙科技股份有限公司)。

## 2 方法与结果

### 2.1 颗粒在流化床中的干燥过程

在流化床锅体的下半部分安装蓝宝石窗口,并将 Thermo Antaris MX 近红外光谱仪的探头固定于窗口上,取样口在流化床的另一侧。将制粒后的湿颗粒倒入流化床中进行干燥,实验所用样品是在流化床干燥过程中从取样口实时取出得到的样品:流化床开始进风干燥并开始连续采集样品光谱,光谱采集完成后立即取样。因物料温度在 30 °C 前样品水分变化较快,故取样间隔时间较短,每隔 30 s 取一次样品;在物料温度高于 35 °C 时水分变化较慢,取样的间隔时间延长,每隔 2 min 取一次样品,待物料温度继续上升到 55 °C 时停止干燥(覆盖了工艺范围以得到全面的样品光谱信息)。一批物料需分成 4 锅流化床完成,每做一锅流化床采集约 10 个样本,一批可采集约 40 个样品,共 12 批采集 519 个样品建立模型。

### 2.2 水分参考值的测定

用水分测定仪测定样品的水分质量含量为 1.00%~7.50%(涵盖了工艺优化值 1.00%~4.00%)。

### 2.3 样品的光谱采集

使用近红外光谱仪以漫反射光纤探头的采集方式连续采集干燥过程中的样品的近红外光谱。光谱的扫描范围为 10 000~4 500  $\text{cm}^{-1}$ ,扫描次数为 64 次,分辨率为 8  $\text{cm}^{-1}$ ,增益值为 8 $\times$ ,每次采集光谱前均进行背景光谱的采集,采集得到的近红外光谱如图 1 所示。

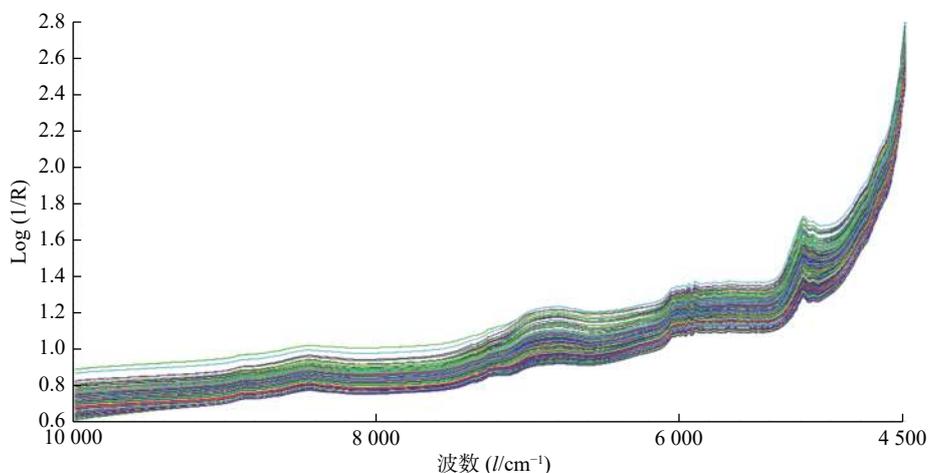


图 1 硫酸羟氯喹颗粒近红外光谱对数图

### 2.4 定量模型的建立

利用 TQ Analyst 9.5 软件结合化学计量学的

PLS 算法,建立硫酸羟氯喹颗粒含水量的定量分析模型。首先采用 Chauvenet 准则检验结合杠杆值-

学生化残差图鉴别并剔除异常样本<sup>[14]</sup>,进而采用 SPXY 分类算法将其余样本划分为校正集和验证集。通过一系列参数,包括校正集误差均方根(RMSEC)、验证集误差均方根(RMSEP)和相关系数  $R$  等评价模型进行考察<sup>[15]</sup>。当模型相关系数  $R$  越接近 1,说明模型拟合效果好,分析准确度越高。当 RMSEC 和 RMSEP 值越小且越彼此接近,即 $|RMSEP-RMSEC|$ 趋近 0,表明模型具有好的稳定性,且当 RMSEP 值较小时,模型具有较高预测能力。

#### 2.4.1 异常样本的剔除

异常样本会对 NIR 模型产生很大的影响,例如误导光谱变量的选择,影响模型的参数估计,降低模型的预测准确度和稳定性<sup>[16-18]</sup>。采用 Chauvenet 准则检验结合杠杆值-学生化残差图,鉴别并剔除 13 个异常样本: 23、58、61、70、79、95、96、97、139、223、276、350 和 473。

#### 2.4.2 样本集划分

在 NIR 模型建立的过程中如何挑选具有代表性的样本,对最终模型的预测准确性与稳定性有很大影响。SPXY<sup>[19]</sup>分类方法是近红外样本分类中一种常用的方法,同时考虑了样本的光谱和浓度特征。进行模型建立的样本共有 519 份,在“2.4.1”项下剔除了 13 份样品,将剩余 506 份样品使用 SPXY 分类法进行分类,344 份样品作为校正集,162 份样品作为验证集。主成分分析结果表明,样本的验证集均匀分布在校正集中。

#### 2.4.3 光谱预处理

光谱预处理方法包括 MSC、SNV 及求导等一系列算法。MSC 通过数学方法将光谱中的散射信号与化学信息进行分离,用于消除由于样品颗粒分布不均及颗粒大小不同所产生的散射对其光谱的影响。SNV 的作用与 MSC 基本相同,均是用于消除由于颗粒散射及光程差异给光谱带来的影响<sup>[15]</sup>。

对光谱进行求导,同时运用 Savitzky-Golay(S-G)平滑滤波,或者 Karl Norris 平滑滤波对光谱进行平滑。求导可以增强光谱信号,平滑可以滤除噪声。从表 1 可以看出 NIR 光谱经过 MSC、一阶导数以及 Karl Norris 平滑,能得到较好的数据处理结果。

#### 2.4.4 建模波段的选择

选择合适的波段用于模型的建立,对最终所建模型的预测准确度和稳定性有重要的影响。为了得到 NIR 光谱的特征信息,首先对前 3 个主成分的载荷向量进行分析<sup>[20]</sup>,图 2A 为前 3 个主成分全光谱范围的载荷图。由 PC1 载荷图可知,光谱在  $4\ 935 \sim 5\ 336\ \text{cm}^{-1}$  和  $6\ 911 \sim 7\ 297\ \text{cm}^{-1}$  附近的波

表 1 不同光谱预处理方法得到的模型参数

预处理方法	RMSEC (%)	RMSEP (%)	RMSECV (%)	$R$
无预处理方法	0.457	0.499	0.489	0.9206
MSC <sup>1</sup>	0.467	0.409	0.481	0.9309
SNV <sup>2</sup>	0.469	0.411	0.483	0.9306
MSC+一阶导数	0.537	0.506	0.585	0.8962
MSC+二阶导数	0.457	0.745	0.867	0.7647
MSC+一阶导数+Norris <sup>3</sup>	0.408	0.435	0.446	0.9372
SNV+一阶导数	0.537	0.506	0.585	0.8962
SNV+二阶导数	0.456	0.745	0.867	0.7647
SNV+一阶导数+Norris	0.405	0.436	0.446	0.9372

注: <sup>1</sup>表示多元散射校正; <sup>2</sup>表示标准归一化变换; <sup>3</sup>表示平滑滤波。

段范围包含更多的特征信息。由 PC2 和 PC3 载荷图可知,在上述波段附近也包含较多的光谱信息。同时,与得到求导后的一阶光谱(图 2B)相比较,确定上述两个波段与水分之间确实具有较强的相关性,是 NIR 光谱中水分的主要吸收峰,可用于水分模型的建立。

#### 2.4.5 主因子数的选择

主因子数的选择在建立模型的过程中至关重要,选择正确的主因子数既能完全利用 NIR 光谱信息,还能避免出现过拟合现象。根据 TQAnlyst9.5 中的 PRESS 图,以 RMSECV 和 PRESS 最小时所对应的主因子数即为最佳主因子数(本研究中主因子数为 6)。

#### 2.4.6 NIR 模型的建立

根据 SPXY 分类法得到的校正集建立模型,用验证集对模型进行验证。以 MSC+一阶导数+Karl Norris 平滑为光谱的预处理方式,选择的建模波段为  $4\ 935 \sim 5\ 336\ \text{cm}^{-1}$  和  $6\ 911 \sim 7\ 297\ \text{cm}^{-1}$ ,结合化学计量学中的 PLS 算法建立 NIR 定量分析模型(图 3):RMSEC 为 0.408,相关系数为 0.9529;RMSEP 为 0.435,相关系数为 0.9366;主因子数为 6。从图 3 可以看出校正集和验证集的数据点在模型范围(1.00%~7.50%)内分布均匀,且校正集分布在验证集的范围,表明用校正集建立的模型对验证集的验证有效,同时 $|RMSEP-RMSEC|$ 的数值为 0.027,趋近于 0,以上结果均表明所建模型较为稳定。

#### 2.4.7 模型准确度的评价

RPD 值是验证集标准偏差与预测集标准偏差的比值,即  $RPD=SDV/SEP$ 。本研究以 RPD 值对模型进行评价。其中,SDV 为验证集所有样本浓度值的标准偏差,SEP 为预测集标准偏差。验证集样

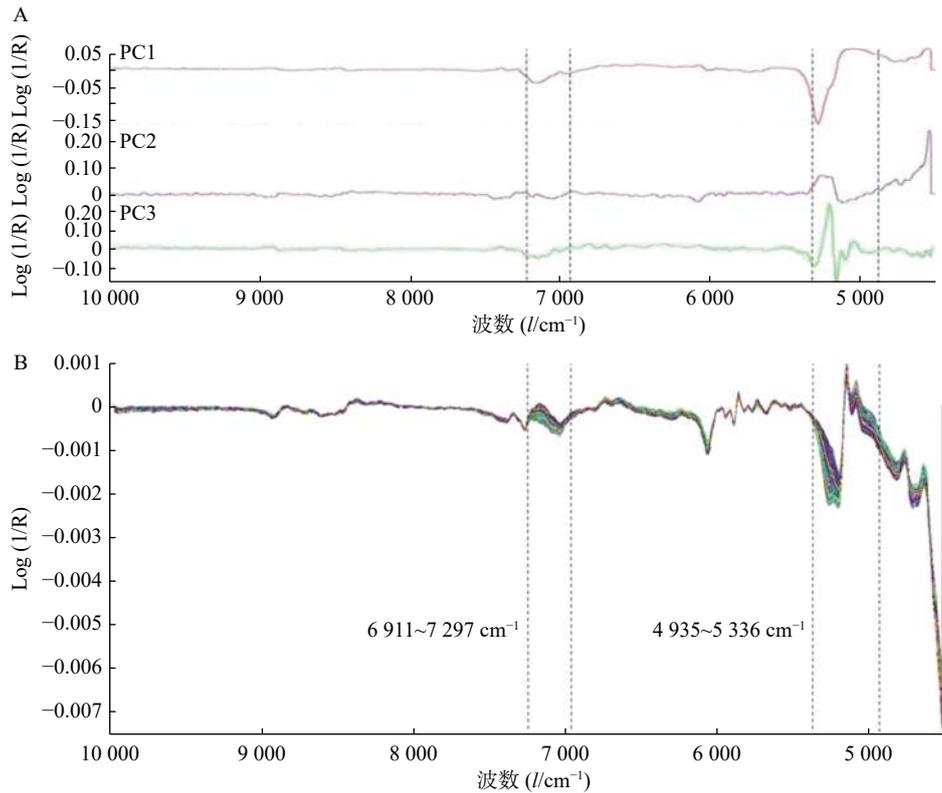


图2 硫酸羟氯喹近红外光谱图

A. 硫酸羟氯喹颗粒前3个主成分载荷图; B. 一阶导数光谱图

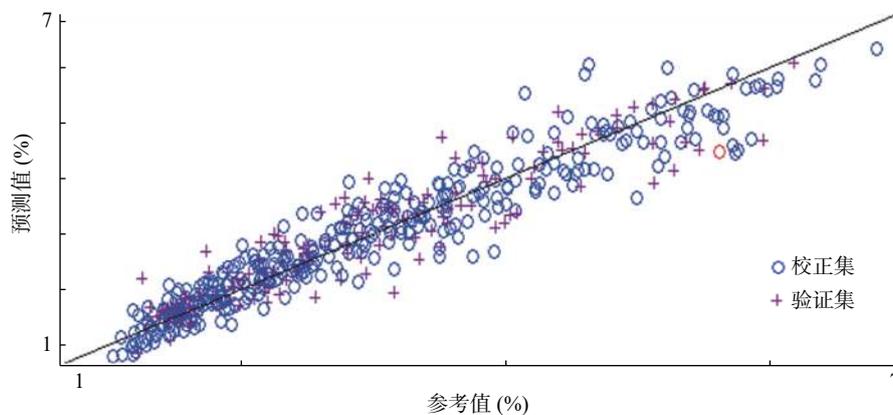


图3 硫酸羟氯喹颗粒定量模型

本的性质分布越均匀, SEP 值越小, RPD 值越大。本研究中, 通过计算得到 RPD 值为 5.18 (大于 5), 表明模型的预测结果可以接受<sup>[10]</sup>。

### 2.5 样品测定

将建立完成的模型应用于一批物料(PAT190601-6)的干燥过程, 按“2.1”项下操作, 采集所取样品光谱。将采集的光谱导入“2.4.6”项下建立的模型, 得到光谱的预测值, 用快速水分测定仪测定每份样品的参考值。预测集的 RMSEP=0.265。将预测集样本的预测值与参考值进行配对样本  $t$  检验, 显著性水平设为 0.05。检验结果  $|t|=0.195$ , 小于  $t$  双尾临

界值 1.990,  $F=1.10 < 1.46$ , 说明预测值与参考值之间无统计学差异。图 4 为预测值与参考值水分变化曲线, 从图中可以看出两个曲线的重合性较好, 进一步表明模型的误差较小, 稳定性较好(鉴于流化床体积, 一个批次样品需要分为 4 份进入流化床干燥)。

### 2.6 模型的应用

在模型已初步建立完成并应用情况较好的小试基础上, 将仪器移入车间, 开始对大生产过程中物料水分变化进行实时检测。将蓝宝石窗口置于流化床底部, 探头切合于窗口, 取样口于窗口的对

面处。考虑到仪器所处环境的改变可能对光谱产生影响,包括吸收峰的偏移、展宽、吸收强度的非线性变化等<sup>[21]</sup>,需要对模型在车间的可应用性重新评价。

对一批物料(PAT-191 223)进行干燥实验,该批样品需分两批进入流化床,每批采集3~4 g样品后,用水分测定仪测定水分值。预测集的RMSEP=0.265。将预测集样本的预测值与参考值进行配对样本 $t$ 检验,显著性水平设为0.05。检验结果 $|t|=0.265$ ,小于 $t$ 双尾临界值2.110, $F=0.78 < 3.18$ ,说明预测值与参考值之间无统计学差异。

将模型导入 Result Operation 软件,可实时反映物料(PAT-191 223)的水分值变化,由图5可见水分值平稳下降、直至趋于稳定。在干燥过程中,在线取出5份样品,用水分测定仪测量参考值并标于图中(红色为参考值,黑色为预测值)。可见预测值与参考值的重合性较好,表明模型的误差较小且稳定性较好。

### 3 讨论

#### 3.1 模型误差来源及解决办法

模型的误差有两个来源:①探头与取样口间隔较远,降低了采集得到的光谱与样品的空间对应性,对模型的准确度带来影响;②取样口的通道较窄,会造成湿物料堆积,对样品的水分值造成影响。解决方法:①光谱采集完成后立即取样,可以保证光谱与样品的时间对应性,对于提高空间对应性的解决方法是可在窗口附近安装取样口;②及时清理取样口堆积的物料,避免对样品水分值产生影响。

#### 3.2 水分预测模型的建立

工艺要求硫酸羟氯喹颗粒水分值为1.00%~4.00%,模型预测范围为1.00%~7.50%,包含了工艺要求的范围,可应用此模型缩小批间的水分差异。例如,当预测值在2.00%时停止干燥,为工艺的改善和提高提供基础。从图5可以看出,部分时间段的水分值上下波动,而非持续下降,且在接近

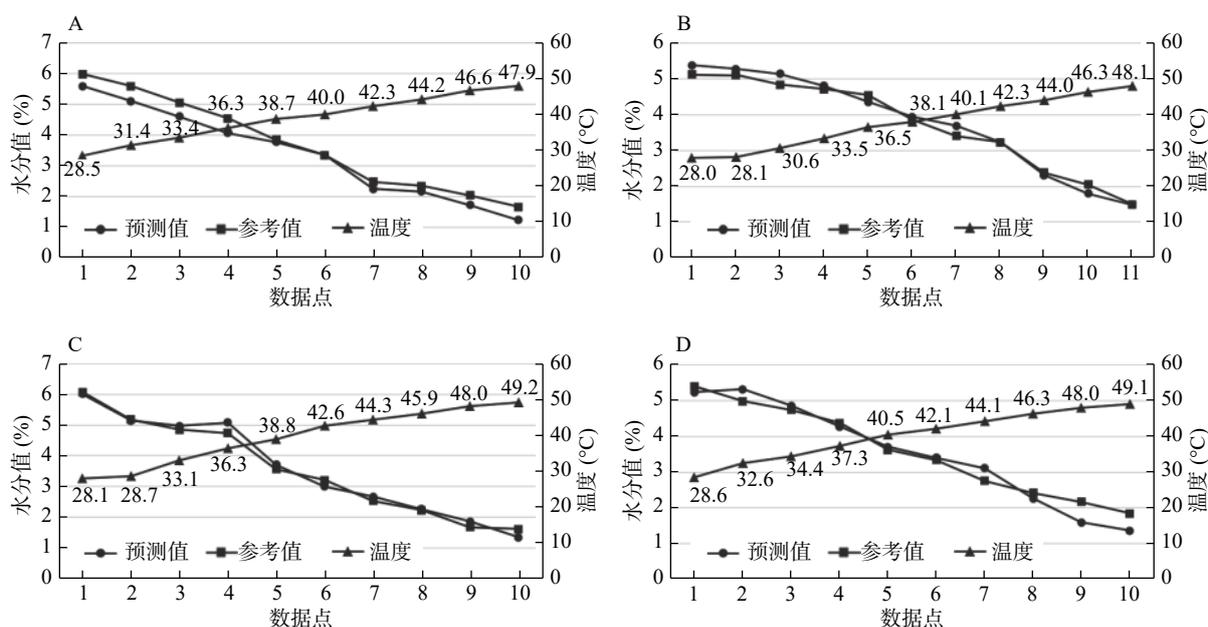


图4 4份样品预测值与参考值的水分变化曲线

A.第1份样品;B.第2份样品;C.第3份样品;D.第4份样品

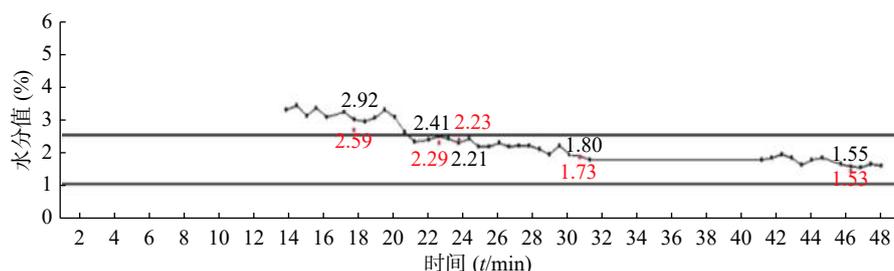


图5 硫酸羟氯喹颗粒水分含量变化图

干燥终点时,水分值在 2.00% 处波动,且无明显下降趋势。原因可能为采集得到的 NIR 光谱反映的是表层样品的光谱,无法准确反映较深处即流化床锅体内部的物料水分情况,而下一张光谱有可能为内部的物料吹到窗口处采得,因此,预测的水分值略有升高。

### 3.3 不同批次样品对水分模型的影响

本实验采集了 11 批次药品用于含水量模型的建立。由于在硫酸羟氯喹的实际生产过程中,原辅料的不同批次之间存在批间差异,此外,生产过程中存在影响硫酸羟氯喹颗粒质量的其他因素,这也导致不同批次的硫酸羟氯喹颗粒之间也存在批间差异。故在建立模型时尽可能包含较多批次,以包含更多批间差异,虽然模型的误差可能会增加,但可提高后期模型在实际大生产中的可应用性。因此,在以后的生产过程中仍需不断累计数据,更新并完善模型。

### 【参考文献】

- [1] FDA. Guidance for industry PAT—A framework for innovative pharmaceutical development, manufacturing and quality assurance. 2004. <https://www.fda.gov/media/71012/download>.
- [2] 褚小立. 近红外光谱分析技术实用手册[M]. 北京: 机械工业出版社. 2016: 127-129.
- [3] FINDLAY W P, PECK G R, MORRIS K R. Determination of fluidized bed granulation end point using near-infrared spectroscopy and phenomenological analysis[J]. *J Pharm Sci*, 2005, 94(3): 604-612.
- [4] MOLTGEN C V, PUCHERT T, MENEZES J C, et al. A novel in-line NIR spectroscopy application for the monitoring of tablet film coating in an industrial scale process[J]. *Talanta*, 2012, 92: 26-37.
- [5] KARANDE A D, HENG P W, LIEW C V. In-line quantification of micronized drug and excipients in tablets by near infrared (NIR) spectroscopy: Real time monitoring of tableting process[J]. *Int J Pharm*, 2010, 396(1-2): 63-74.
- [6] 年秀玲. 类风湿性关节炎应用硫酸羟氯喹治疗的安全性和疗效探析[J]. *北方药学*, 2016, 13(12): 145.
- [7] 王文灵. 硫酸羟氯喹治疗风湿病的临床效果研究[J]. *中国现代药物应用*, 2016, 10(12): 138-139.
- [8] 魏薇, 何金昌, 余素君. 羟氯喹治疗系统性红斑狼疮的效果评价[J]. *当代医药论丛*, 2017, 15(16): 138-139.
- [9] YAMAMOTO Y, SUZUKI T, MATSUMOTO M, et al. Evaluation of the degree of mixing of combinations of dry syrup, powder, and fine granule products in consideration of particle size distribution using near infrared spectrometry[J]. *Chem Pharm Bull*, 2012, 60(5): 624-631.
- [10] 李沙沙, 陈辉, 赵云丽, 等. 硫酸羟氯喹颗粒水分含量测定近红外定量模型的建立[J]. *沈阳药科大学学报*, 2019, 36(7): 593-599.
- [11] 王维皓, 张永欣, 冯伟红, 等. 基于近红外光谱及组合间隔偏最小二乘法的天南星中水分及总黄酮含量测定研究[J]. *中国中医药信息杂志*, 2017, 24(7): 72-75.
- [12] 王遥琼, 王维皓. 近红外光谱用于甘草中甘草苷、甘草酸及水分测定[J]. *中国实验方剂学杂志*, 2016, 22(19): 59-62.
- [13] 张翠, 柴欣生. 近红外光谱法快速测定甘蔗渣水分和糖分含量[J]. *中华纸业*, 2011, 32(22): 55-58.
- [14] CHEN Y, XIE M Y, ZHANG H, et al. Quantification of total polysaccharides and triterpenoids in *Ganoderma lucidum* and *Ganoderma atrum* by near infrared spectroscopy and chemometrics[J]. *Food Chem*, 2012, 135(1): 268-275.
- [15] 杜敏. 中药近红外光谱检测影响因素的研究[D]. 北京中医药大学, 2013.
- [16] FERNANDEZ PIERNA J A, WAHL F, de NOORD O E, et al. Methods for outlier detection in prediction[J]. *Chem Intell Lab Sys*, 2002, 63(1): 27-39.
- [17] 闵顺耕, 李宁, 张明祥. 近红外光谱分析中异常值的判别与定量模型优化[J]. *光谱学与光谱分析*, 2004, 24(10): 1205-1209.
- [18] 刘蓉, 陈文亮, 徐可欣, 等. 奇异点快速检测在牛奶成分近红外光谱测量中的应用[J]. *光谱学与光谱分析*, 2005, 25(2): 207-210.
- [19] GALVAO R K, ARAUJO M C, JOSE G E, et al. A method for calibration and validation subset partitioning[J]. *Talanta*, 2005, 67(4): 736-740.
- [20] 孔雾虹. 青霉素类产品专属在线混合均匀度监测体系设计[D]. 山东大学, 2014.
- [21] 张学博, 冯艳春, 胡昌勤. 近红外多元校正模型传递的进展[J]. *药物分析杂志*, 2009, 29(8): 1390-1399.

【收稿日期】 2020-07-31 【修回日期】 2020-11-19

【本文编辑】 李睿旻