

附件：多变量统计过程控制技术指导原则草案

多变量统计过程控制技术指导原则

本指导原则介绍了基于多变量数据分析和统计的建模过程控制技术在药品制造过程监控中的应用，为药品生产质量管理的规范实践提供技术手段，也为制药行业中的多变量统计过程控制技术提供统一的方法和标准，确保药品生产的安全性和药品质量的一致性。

1 概述

1.1 统计过程控制

统计过程控制（Statistical Process Control, SPC）作为一种过程性能监控和异常诊断的技术工具，是运用数学统计分析技术监视和控制生产过程，从而保证产品质量的有效手段。SPC 通过统计分析生产过程中的变量构建控制图进而实现监测和控制生产过程的目的。SPC 在制药行业中越来越多地被应用于寻求质量管理体系的持续改进，通过根据过程的变量来监控制药过程的稳定性。

1.2 多变量统计过程控制

多变量统计过程控制（Multivariate Statistical Process Control, MSPC）是一种多变量数据分析与统计建模技术，使用统计模型同时监测多个相互关联的变量，这些变量组合在一个测量向量中，变量之间可能高度相关（例如光谱图），也可能不相关。MSPC 结合多变量统计分析方法，实现原始高维数据空间向低维特征空间的投影，得到一组保留原始特征信息且摒弃了冗余信息的特征变量，是分析处理高维数据的有效工具。MSPC 的基本原理为采用偏最小二乘法（Partial Least Squares）、主成分分析（Principal Component Analysis, PCA）等统计分析方法进行数据降维，建立统计模型，通过模型计算得到过程轨迹，根据数据波动范围确定正常区间，将超出该区间的过程判定为异常过程，最终实现生产过程的监视和控制。药品制造中的许多过程都可能应用 MSPC，例如工艺开发、生产运营监控、流程优化、故障排除、故障检测等。

2 MSPC 的实施流程

2.1 数据采集

当开始对过程进行多变量统计过程分析时，变量分布是不明确的，因此必须

收集过程数据以开发模型。多变量统计过程控制的实施需要一个能够反映各过程变量之间内在联系的监控模型，因此首先需要采集过程运行的历史数据来建立这个监控模型。如历史数据不包含异常数据，则应人为改变生产过程的质量参数或工艺参数以生成异常值。数据采集主要依靠过程分析技术，包括质控指标过程分析技术和工艺参数过程分析技术。前者以光谱技术为主，后者主要包括各类传感器如温度传感器、压力传感器等，分别采集得到光谱数据和工艺数据。

2.1.1 光谱数据

拉曼光谱法（通则 0421）和近红外光谱法属于非破坏性光谱学方法，常用于生产过程样品光谱数据的采集和生产过程的在线监测。采集得到的光谱数据可参考化学计量学指导原则进行预处理才能进行后续的数据降维处理。

2.1.2 工艺数据

关键工艺参数（Critical Process Parameter, CPP）是对关键质量属性有影响的过程参数，在生产过程中，应对其进行监测和控制以确保过程输出所要求的质量。CPP 是工艺步骤或单元操作的输入运行参数（速度、流速等）或工艺状态变量（温度、压力等）。实际工作中可以通过对工艺参数的多变量分析来判断影响药品质量的关键因素。例如在中药材提取过程中，药材的粉碎程度、提取溶剂、提取温度、提取时间等因素都会对提取物的提取质量产生一定影响；片剂生产工艺中，碾压时的螺旋钻速度、压片时的压缩速度和强制进料器速度对片剂溶出率有显著影响；抗体药物细胞培养过程中反应器的 pH 值、温度、渗透压、二氧化碳浓度等，基础培养基的保存时间、最终 pH 值和渗透压等对目标产品质量也有一定影响。

2.2 数据处理

采集得到的数据通常是三维数据矩阵 $X(I \times J \times K)$ ：即 I 个正常操作批次（Normal Operation Condition, NOC），在 J 个时间点采集了 K 个变量的数据。首先对此三维数据矩阵作降维处理，通常根据实际需求选择适宜的降维方式，按变量方向展开和按批次方向展开为常用的两种方式。若按变量方向展开，则得到二维矩阵 $X(JI \times K)$ ，即将批次和时间这两个维数组合成为矩阵的列，保留过程变量的维数成为矩阵的行。因此，这个二维矩阵的列包含相对应过程变量在所有批

次和所有时间点上的测量数据，通常用于监控批次内运行并指示期间异常；同样地，若按批次方向展开，则生成二维矩阵 $X(I \times KJ)$ ，能够跟踪随时间变化的动态并监控过程参数和批次端点属性之间的协方差。

得到的二维数据矩阵需要在每一采样时刻对过程变量进行均值中心化以及方差归一化的标准化处理，以消除不同批次之间存在的大部分动态特性及非线性误差，也称为轨迹的同步化。例如对于矩阵 $X(JI \times K)$ ，计算公式如下：

$$\tilde{X}_{ji,k} = \frac{X_{ji,k} - \overline{X_k}}{S_k}$$

其中， $X_{ji,k}$ 为矩阵 $X(JI \times K)$ 的第 $j \times i$ 行、第 k 列数据， $\overline{X_k}$ 为矩阵第 k 列的均值， S_k 为矩阵第 k 列的标准偏差。

2.3 过程数据的主成分分析

在药物生产过程中可能有多个作为变量被测量的工艺参数和质量参数，然而，实际影响最终产品质量的关键参数可能仅是所有被测量参数中的少数几个。主成分分析（Principal Component Analysis, PCA）可以将高维的多变量或者多元的数据样本，通过投影和线性组合的方式降维到由少数变量表示的低维空间，从大量数据中提取反应过程运行的有效信息，即可建立变量随时间有序变化的统计控制模型图。PCA 的作用是找出一组反映这些特征的新变量，后续针对这组变量进行的检验分析和控制工作可以大幅提高工作效率。这些新变量被称为主成分（Principal Components）或潜在变量（Latent Variables），它们是原始过程变量的线性组合，最大程度地携带了原始变量的有效信息。

进行 MSPC 控制时，首先要建立 PCA 模型，该模型能够以低维的数据展示对过程影响主要的特征变量从而反映过程的正常运行情况。建立 PCA 模型时，为了便于计算数据和消除数据量纲等带来的潜在影响，需要进行标准化处理，即计算该变量中每个值减去均值后与它的标准差的比值。将正常生产过程中的运行数据定义为 $X \in R^{n \times m}$ ，对 X 按下式进行标准化：

$$\overline{X_s} = [X - (1 \cdots 1 \cdots 1)^T M] \text{diag} \left\{ \frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_m} \right\}$$

式中 $M = [m_1 \ m_2 \ \dots \ m_m]$ 代表 X 变量的平均值， $s = [s_1 \ s_2 \ \dots \ s_m]$ 为变量的标

$$\overline{X} = t_1 p_1^T + t_2 p_2^T + \cdots + t_m p_m^T$$

84 准差。对 $\bar{X}$ 进行主成分分析，可以得到下方公式：

85

86 建立主成分模型之前，需要确定参与主成分模型的主成分个数，使用前
87 $k(k < m)$ 个主成分数，可得主成分模型如下：

88
$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E = X_p + E$$

89 其中 E 为偏差，表示数据中的噪声和次要信息。

90 对过程数据进行主成分分析，基于前两个主成分（Principal Component, PC）
91 构建得分散点图，可以反映过程数据 PC1、PC2 得分随时间变化的趋势，正常操
92 作批次（NOC）的数据变化趋势应基本一致。

93 2.4 MSPC 模型的建立

94 在建立 MSPC 模型时，常采用过程数据的主成分得分（PC Score）、Hotelling's
95 T^2 和 DModX 等统计量对过程进行描述。基于正常批次（NOC）的过程数据可以
96 得到每种统计量的过程运行轨迹和控制限，进而构建过程控制图。以某提取过程

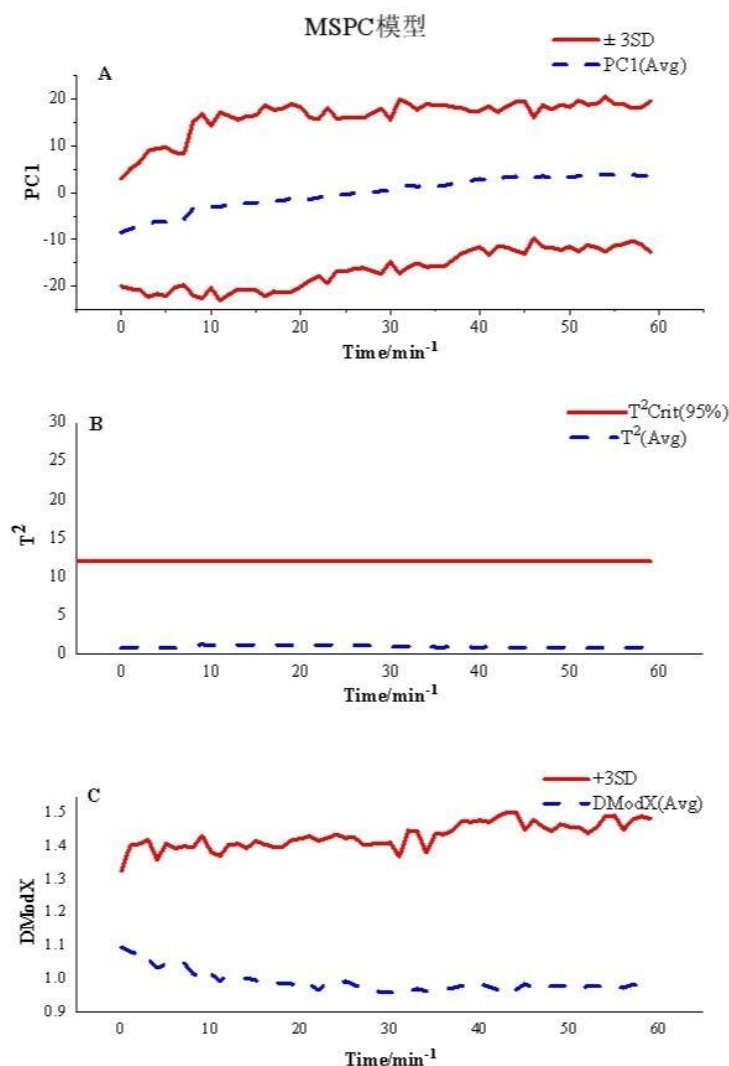


图 1 某提取过程 MSPC 模型 (A 第一主成分得分控制图; B Hotelling's T^2 控制图; C DmodX 控制图)

MSPC 模型为例, 图 1 为第 1 个主成分得分, Hotelling's T^2 和 DmodX 统计量控制图。红色实线代表控制限, 蓝色虚线代表多批次该统计量的平均值。

PC1 得分控制图采用 $\pm 3SD$ 作为控制图的上下限; Hotelling's T^2 控制图采用 95%的控制限; DModX 控制图, 采用+3SD 作为控制图的上限。多变量控制图可用于同时监控存在着关联关系的两个或多个质量特性, 在实际运用过程中, 往往需要综合分析两种或两种以上的控制图。多变量控制图的开发可以通过 SIMCA-P+、Unscrambler、Matlab 等软件实现。

主成分得分轨迹反映的是各主成分随时间的变化趋势, 可直观地表征过程变

化趋势。实验过程中常利用第一主成分对过程进行监控；Hotelling's T^2 统计量指示每个采样信息在变化趋势和幅值上偏离主成分模型的程度，表征 PCA 模型内部的变换情况，反映变量的变动是否异常；DModX 表示采样数据在自变量 X 空间到主成分模型的距离，主要通过主成分模型外部残差矩阵来反映过程变化的情况，为模型的外部数据变化度量。相关统计量的具体描述如下：

2.4.1 Hotelling's T^2 统计量

T^2 统计量表示采样数据在主元空间上投影的变化情况，指示每个采样信息在变化趋势和幅值上偏离主成分模型的程度，表征的是 PCA 模型内部的变换情况。某个批次在时间点 k 处的 T^2 值为得分向量 t_k 的标准平方和，计算公式如下：

$$T_k^2 = T_k \Lambda_R^{-1} T_k^T = \sum_{r=1}^R \frac{t_{kr}^2}{\lambda_r}$$

式中， T_k 为 k 时刻 NIR 光谱的得分向量； Λ_R 为前 R 个主成分对应的特征值组成的对角矩阵； R 为 PCA 模型的主成分数； t_{kr}^2 为得分向量 T_k 在第 r 个主成分上的得分值； λ_r 为对角阵 Λ_R 中第 r 个主成分对应的特征值。

Hotelling's T^2 控制限可以利用 F 分布来计算，公式如下：

$$T_{UCL}^2 = \frac{R(I-1)}{I-R} \bullet F_{\alpha}(R, I-R)$$

式中， R 为主成分个数， I 为训练集批次数， α 为置信度， $F_{\alpha}(R, I-R)$ 为置信度 α 下第一第二自由度为 $(R, I-R)$ 时 F 分布的临界值。若某生产批次在 k 时刻计算得到的 $T_k^2 \leq T_{UCL}^2$ ，说明该时刻过程处于正常状态。

2.4.2 DModX 统计量

X 区块模型距离（Distance to the Model X，DModX）表示样本数据在自变量 X 空间到主成分模型的距离即采样点偏离模型的绝对距离，表征主成分模型不能解释的部分，反应的是主成分模型的外部数据变化的大小，主要用来监测输入的数据结构是否发生变化。某个批次在 k 时刻的 DModX 值计算公式如下：

$$DModX_k = \sqrt{\frac{\sum_{j=1}^J e_{kj}^2}{J-R}}$$

式中, J 为光谱变量数; R 为主成分数; e_{kj} 为 k 时刻光谱在变量 j 上的残差, 计算公式如下:

$$e_{kj} = x_{kj} - \tilde{x}_{kj}$$

式中, x_{kj} 为 k 时刻光谱在变量 j 上的原始值; $\tilde{x}_{kj}$ 为 k 时刻光谱在变量 j 上的得分值。为了保证不同时间点样本间数据的整齐可比性, 需要对得到的 DModX 值进行如下标准化处理:

$$(DModX, N)_k = \frac{DModX_k}{\bar{D}_{cal}}$$

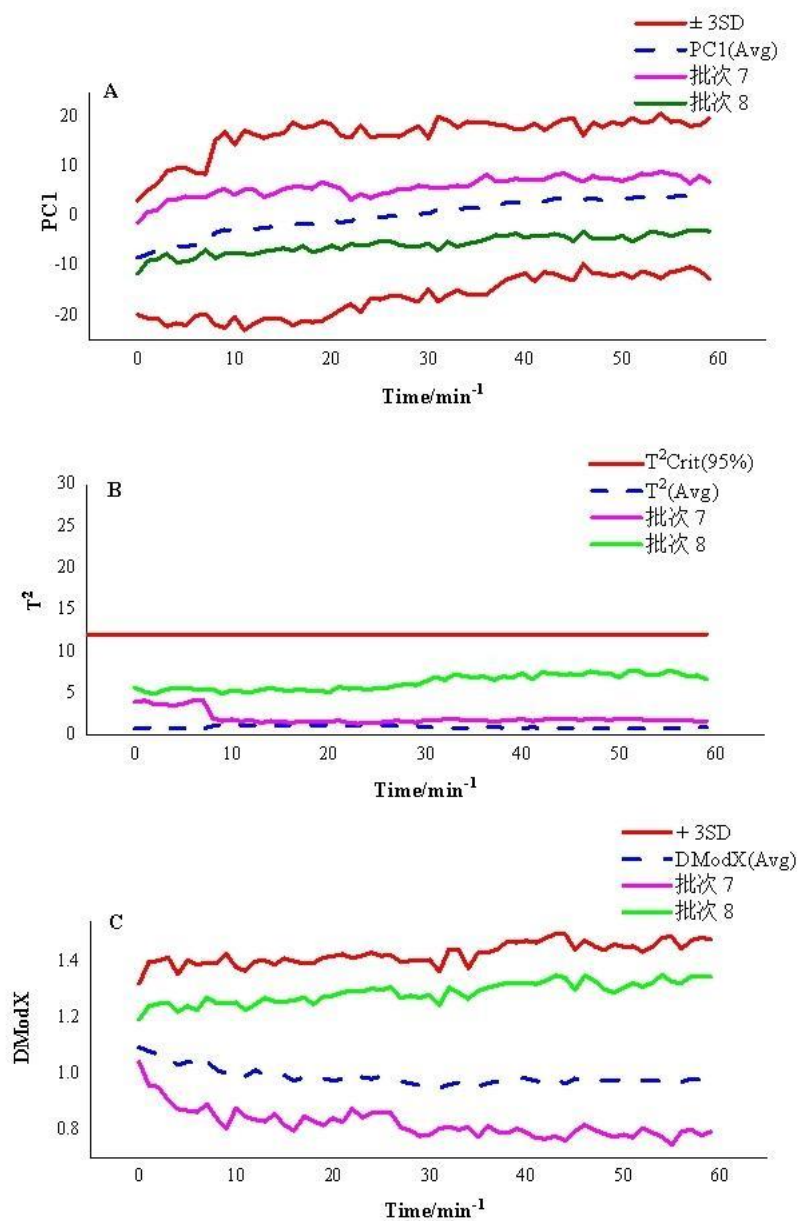
DModX 控制限的计算公式如下:

$$DModX_{lim} = \bar{D}_{cal} + 3SD$$

式中, $(DModX, N)_k$ 为 k 时刻样本 DModX 的标准化值, $\bar{D}_{cal}$ 为训练集的 DModX 均值, SD 为标准偏差。若 $(DModX, N)_k \leq DModX_{lim}$, 表示此时此刻的 DModX 统计量正常, 反之, 则说明当前时刻过程出现异常。

2.5 MSPC 模型的验证与应用

MSPC 模型的验证与应用, 即测试批次的多变量控制图的开发, 通常依次采用未构建模型的正常批次 (NOC) 验证和异常批次 (AOC) 验证, 基于随机选取的用于批次验证的 NOC 组和人为干预形成的异常组 (模拟过程的异常工况) 数据绘制控制图。评价 MSPC 模型对工艺过程的监测能力, 得到测试批次的多变量控制图。如图 2 所示, 所建立的模型性能良好, 无论是 PC1 得分控制图还是 Hotelling's T^2 、DModX 控制图, 正常批次都能落在控制模型范围内, 表明批次过程整体运行正常即所获得的批次质量在可接受的波动范围之内。如图 3 所示, 异常批次则至少落在其中一种控制图的波动范围之外。在利用 MSPC 模型在线监控生产过程时, 需联合使用主成分得分、Hotelling's T^2 和 DModX 统计量 3 种控制图, 当 3 种统计量均在控制限以内时则判断此批次处于受控状态, 其余则认定为异常情况。



160

161

162 图 2 正常批次 7-8 监控图 (A 第一主成分得分控制图; B Hotelling's T^2 控制图;

163 C DmodX 控制图)

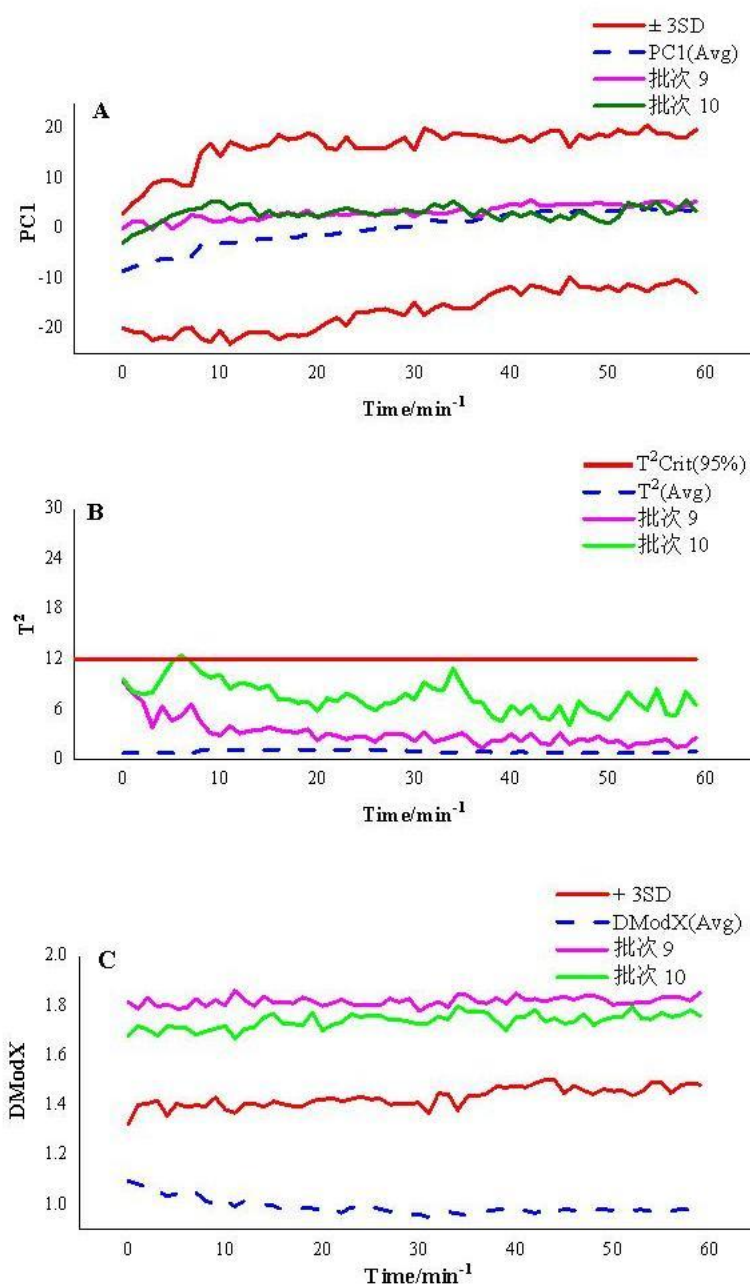


图 3 异常批次 9-10 监控图 (A 第一主成分得分控制图; B Hotelling's T^2 控制图; C DmodX 控制图)

2.6 MSPC 模型的维护

利用已经建立的模型对未知过程数据进行预测, 考察 PC1 得分、Hotelling's T^2 及 DModX 控制图是否出现异常点, 对异常产生原因进行判定。若异常点为缓慢逐渐出现的情况, 需要更新模型及控制图。

起草单位：浙江大学药学院
参与单位：温州大学生命与环境科学学院等
主要起草人及联系电话：陈勇，0571-88208621

多变量统计过程控制技术指导原则起草说明

本指导原则参考国内外相关标准和技术规范，结合我国制药工业应用特点制定，为医药行业中的多变量统计过程控制技术提供统一的方法和标准，并进一步优化补充相关质量管理技术要求，促进多变量统计过程技术在我国制药行业的应用，从而更好地加强药品质量控制，促进药品高质量发展。本指导原则主要包括概述和多变量统计过程控制的实施流程，描述了算法原理、数据采集、数据处理、过程数据的主成分分析和多变量统计过程控制模型的建立等。

公示稿