

土霉素发酵生产后期模型化

徐川育¹ 方崇智¹ 马振华²

(清华大学自动化系¹和应用数学系², 北京 100084)

本文针对间歇式土霉素发酵过程(FBOFP)中生产期后期产物增加慢的较普遍问题,采用了两阶段最小二乘法和 Durbin 程序,建立培养因子,生化因子和产物因子的多对多回归模型,得出培养基残氮低是造成比生产率低的结论。但原有工艺却认为残氮是不低的,其原因在于以前的氮测量法不能把有用氮与无用氮区分开来,采用了新引进的测量法后,结果发现:无用氮不低,而有用氮却是显著的时间减函数,从而证实了本文的结论是完全正确的,3000h工业试验的结果表明:产量在原来较高基础上增加1%,发酵效价达到32000 μ /ml。

关键词 生产期后期;多对多回归模型;残氮;测量

在采用非流加中间补料工艺的土霉素发酵过程中,存在一个问题:其前期通过人工代谢调控虽然可以使发酵效价有较大提高,但到后期,发酵效价增长缓慢甚至停止。现在较多地归因于培养基粘度大的外观原因,而认为现有工艺是符合菌丝后期生长条件的,现工艺要求 pH 和残氮随时间增长而大幅度递减。在模型化方面,应用平衡方程来研究氮代谢还存在困难,表现在随着菌令增长,菌丝衰亡,菌浓减少和无法实时检测,该问题的解决未见报道,本文在工业试验数据的时间序列分析方面进行研究。

材料和方法

(一) 材料

1. 菌种: *Streptomyces armillatus* 138, 由华北制药厂提供。

2. 培养基组成(%): 黄豆饼粉3.0, 玉米浆0.65, 淀粉8.0, $(\text{NH}_4)_2\text{SO}_4$ 1.2, NaCl 0.2, CaCO_3 1.1, KH_2PO_4 0.015, 每罐外加 CoCl_2 500g。

3. 发酵罐: 50m³ 发酵罐16个。

(二) 方法

1. 培养条件: 残糖40—50mg/ml, 氮氮0.1mg/ml, 空气流量1900m³/h, 罐温30 $^{\circ}$ C, 搅拌转速140r/min。

2. 采样和补料: 每八小时采样生化测定和补料各一次, 补料原则为: 到下次采样前残糖和残氮分别不低于下限 30 mg/ml和0.1mg/ml。

模型化方法

模型化过程如下: 由试验批报, 获得培养因子(补糖量和通氮量), 生化因子(残氮, 残糖和pH)和生物因子(效价和总亿)(总亿为每罐抗生素总产量)的观测值, 将所有观测值按批报采样时序排成一个时间序列, 以自相关系数为指标, 通过差分, 可以得到一个平稳列。根据机理, 作出模型结构图, 按图写出联立方程组后进行参数估计。为避免不一致性和偏倚

本文于1992年1月20日收到。

国家七五重点科技攻关项目“75-53-01-13/01”。

性，宜采用两阶段最小二乘法，参数得到后还必须再行校正，以消除序列相关，可采用 Durbin 程序，最后对得到的模型参数还必须进行统计假设检验，包括 F-检验和 DW-检验。

建模流程见图 1。

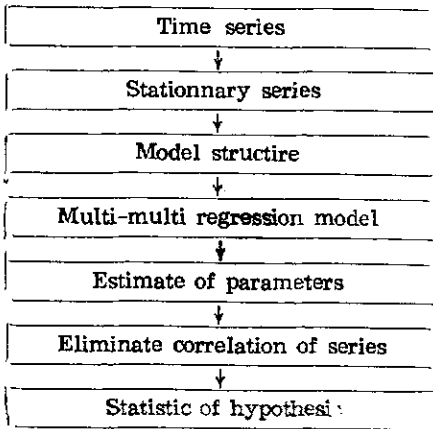


图 1 建模流程

Fig.1 Flowchat of modelling

(一) 有关概念和方法

定义 1. 联立方程组，由一组若干个方程组成，每个方程的参数由 OLS 得出，

引理 1. 联立方程组的估计参数是不一致的，偏倚的。

定义 2. 内生变量：由联立方程系统决定，不仅影响系统也受系统影响，具有某种概率分布的随机变量，它的参数是联立方程组系统估计的元素。

定义 3. 外生变量：或者是已知数的集合，即没有概率分布的固定变量；或者是具有临界概率分布的随机变量，其参数不是系统研究的元素，这些变量由方程组以外来决定。

定义 4. 多对多回归模型：一个描述变量关系结构的完整的方程系统，模型形式为每个内生变量表示成其他内生变量，外生变量和扰动项的函数。

定义 5. 简化式模型：多对多回归模

型中的内生变量仅为外生变量和随机扰动项的函数。

(二) 两阶段最小二乘法和 Durbin 程序用于模型化工作

1. 两阶段最小二乘法^[1]：第一阶段：假定多对多回归模型为：

$$Y_i = \beta_{i1}Y_1 + \beta_{i2}Y_2 + \dots + \beta_{im}Y_m + \gamma_{i1}X_1 + \dots + \gamma_{ik}X_k + U_i - \beta_{ii}Y_i \quad (1)$$

其中 $Y_i, i=1, \dots, m$ 为内生变量；

$X_j, j=1, \dots, k$ 为外生变量；

U_i 为扰动项。

把多对多回归模型化为下列简化式模型：

$$Y_i = \pi_{i1}X_1 + \pi_{i2}X_2 + \dots + \pi_{ik}X_k + V_i \quad (2)$$

其中 $\pi_{ij}, i=1, \dots, m, j=1, \dots, k$

对样本观测值 (X_1, \dots, X_k, y_i) 用 OLS 进行简化式方程的参数估计，得到估计方程为：

$$\hat{Y}_i = \hat{\pi}_{i1}X_1 + \dots + \hat{\pi}_{ik}X_k \quad (3)$$

用观测值 (x_1, \dots, x_k) 代入 (3) 式得到 Y_i 的估计值 \hat{Y}_i ，于是方程的误差项为：

$$V_i = y_i - \hat{y}_i \quad (4)$$

第二阶段：在多对多回归模型 (1) 式中，采用数据样本：

$$(\hat{Y}_1, \dots, \hat{Y}_m, X_1, \dots, X_k) \quad (5)$$

对方程 (1) 式用 OLS 作参数估计，得 $\hat{\beta}_{ii}$ 和 $\hat{\gamma}_{ij}, i=1, \dots, m; j=1, \dots, k; i=1, \dots, m$

2. Durbin 程序(广义差分法)^[2]：

(1) 不失一般性，设方程组 (1) 的每一个方程在 t 时刻有如下形式：

$$\zeta_t = \alpha_0 + \alpha_1\eta_{1,t} + \dots + \alpha_{m+k-1}\eta_{m+k-1,t} + \varepsilon_t \quad (6)$$

用 OLS 估计 $\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}$ ，求一阶序列相关系数 $\hat{\rho}$ ：

$$\hat{\rho} = \frac{cov(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})}{\hat{\sigma}_\varepsilon^2}$$

$$= \frac{cov(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})}{\sqrt{var(\hat{\varepsilon}_t)var(\hat{\varepsilon}_{t-1})}} \quad (7)$$

(2) 方程组(3)变为:

$$\xi_t - \hat{\rho} \xi_{t-1} = a_1(1 - \hat{\rho})$$

$$+ a_2(\eta_{2,t-1} - \hat{\rho}\eta_{2,t-1})$$

$$+ \dots + a_{m+k-1}(\eta_{m+k-1,t} - \hat{\rho}\eta_{m+k-1,t-1}) + \varepsilon_t \quad (8)$$

用OLS对上式估计 $a_k, n=1, \dots, m+k-1$

(3) 比较式(1)和式(2)得出两个DW值, 验证 $\hat{\varepsilon}_t$ 与 $\hat{\varepsilon}_{t-1}$ 的序列相关是否有所减少。

3. DW-检验^[3]:

已知 $U_t = \rho U_{t-1} + V_t$

假设 $H_0: \rho = 0$

对立假设 $H_0: \rho \neq 0$

统计量

$$DW = \frac{\sum(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum \hat{\varepsilon}_t^2} \quad (9)$$

结果和讨论

1. 工艺的模型结构见图2:

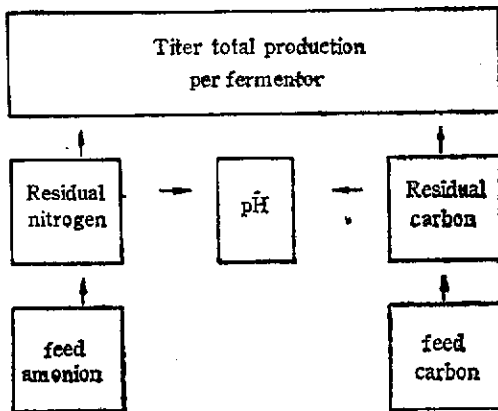


图2 工艺模型结构

Fig.2 Model structure of artificial culture

模型如下:

$$\mu(t) = 0.97 + 0.22pH(t)$$

$$+ 1.02r_N(t) - 0.38r_C(t) \quad (11)$$

$$F = 14.7, R^2 = 0.8,$$

$$DW = 1.98, \rho = 0.01$$

$$r_C(t) = 1.05 - 0.30r_N(t)$$

$$+ 0.14F_C(t-1) \quad (12)$$

$$F = 17.86, R^2 = 0.96,$$

$$DW = 2.05, \rho = -0.04$$

$$r_N(t) = 4.32 - 2.46r_C(t)$$

$$+ 0.14F_N(t-1) \quad (13)$$

$$F = 24.39, R^2 = 0.77,$$

$$DW = 1.94, \rho = 0.02$$

其中 $\mu(t)$ 为 $[t-1, t]$ 区间内效价增量, $F_C(t-1), F_N(t-1)$ 为 $t-1$ 时刻的补糖量和通氮量, $pH(t), r_C(t)$ 和 $r_N(t)$ 分别为 t 时刻的pH值, 残糖量和残氮量, F, DW 分别为关于参数估计的F-检验和DW-检验统计量的计算值, R^2 为线性相关系数, ρ 为序列相关系数。

以生化解释上述模型表明有如下氮代谢方面的规律: (1)若通氮量大, 会引起基质残氮高, 则能使效价增量; (2)若pH高, 则效价增量, 欲使pH高, 又必须多通氨水。

但是, 上述结果与原有测量结果发生矛盾, 原测量结果为基质中残氮已经够了, 故工艺要求减少通氮量, 但是, 从机理上看, 在生产期后期, 部分菌丝自溶和培养液中废料增加, 导致氮浓度增加, 这些氮是不能利用的, 应将它和有用氮区分开来, 可是原有的测量法不能反映出来, 新引进的测量法表明: 有用氮是时间的减函数, 无用氮是时间的增函数, 新引进的测量结果完全证实了本文的结论。

2. 进行了加大后期通氮量的工业试验: (1)试验罐批: 百吨大罐试验共十七个罐批, 对照罐十六个, 每罐批时间190

小时。

(2) 试验与对照数据: 见表 1。

表 1 对照试验数据
Table 1 Experimental result

	Experiment	Contrast
Standard deviation of production conc. (μ/ml)	769,666	1265,251
Expectation of total production per fermentor (μ/ml)	15351,505	15274,306
Standard deviation of total production (μ/ml)	371,757	611,393

(3) 统计假设检验。假设 H_0 :

$$\sigma_{\text{实验}} = \sigma_{\text{对照}}$$

$$n_1 = 16, S_{\text{对照}} = 611.39,$$

$$n_2 = 17, S_{\text{试验}} = 371.75$$

σ 为方差, 其中 n_i 为自由度, $i = 1, 2$, S 为标准差。

$$F = \frac{S_{\text{试验}}^2}{S_{\text{对照}}^2} = \frac{(611.39)^2}{(371.75)^2}$$

$$F \sim F(f_1 = 15, f_2 = 16)$$

当 $\alpha = 0.1$ 时,

$$\lambda = F_{0.5\alpha} = F_{0.05}(15, 16) = 2.4$$

$$\therefore F > \lambda$$

故否定原假设, 认为两个总体方差有显著差异。

(4) 产量提高情况: 概率分布为

$$P_{\text{试验}} = 0.5, P_{\text{对照}} = 0.49$$

$$P_{\text{试验}} - P_{\text{对照}} = 0.01$$

参 考 文 献

- [1] 张 寿, 于清文: 计量经济学, 上海交大出版社, pp. 176—183, 193—195, 243—247, 260—262, 1984年。
 [2] 李卓立: 实用经济计量模型与经济预测, 清华大学出版社, pp. 25—27, 1981年。
 [3] Draper, N.R., Smith, H.: Applied Regression Analysis, John Wiley Sons, Inc. USA, pp. 162—166, 1981.

Oxytetracycline Fermentation Later Phase Modelling

Xu Chuangyu Fang Chongzhi Ma Zhenhua

(Department of Automation and Application Chemistry, Tsinghua University, Beijing 100084)

This paper investigates the low rate of the production in the late production of FBOFP (Fed-Batch Oxytetracycline Fermentation Processes), suggests a multi-to-multi regression model, and results that the low concentration of the residual nitrogen in substrate decreases the rate of production, by means of two-stages OLS and Durbin's procedure.

But the former artificial culture considered the concentration of the residual nitrogen is not low. The original measurement did not divide available nitrogen from non-available nitrogen. The newly imported measurement shows that the unuseful nitrogen in the substrate increasing but the useful nitrogen decreasing as time increasing. This proves that

the result of this paper is completely correct.

The result of the industrial experiments in large scale shows that the output is 1% more than the original output per year, and the titer reaches 32000 μ /ml.

Key words Late production phase; residual nitrogen; measurement; structural model