

## 洁霉素发酵平衡生产期的糖耗速率预测

徐川育<sup>1</sup> 方崇智<sup>1</sup> 马振华<sup>2</sup>

(清华大学 自动化系<sup>1</sup>和应用数学系<sup>2</sup>, 北京 100084)

本文研究了工业规模的批式洁霉素发酵过程中平衡生产期中有效的糖耗速率预测模型,在我国洁霉素批式发酵过程中首次设计并实现了流加补料的计算机系统。洁霉素产量可提高5%。

**关键词** 批式发酵; 糖耗速率预测; 流加

补料批式洁霉素发酵过程生产期中有一个较长的平衡生产期,平衡生产期中微生物增殖速率与死亡速率几乎相等,达到动态平衡。此时活细胞总数较多,比生产速率较高。平衡生产期是个拟稳态过程。其产物形成动力学可近似用非生长关联方程(Non-growth associated equation)来描述<sup>[1]</sup>:

$$r_p = \beta X \quad (1)$$

它的意义为平衡生产期的洁霉素比生产速率主要取决于菌量,并近似与菌量成正比。

由于菌体处于半饥饿状态时才能分泌较多的洁霉素,而现在我国洁霉素发酵过程中采用的补料方式几乎都是批式的,致使基质浓度高低不均、菌体不能保持半饥饿状态,和导致产生基质抑制作用。因此,只有进行缓慢连续流加补料才是解决的办法。这一想法提出几十年了,国外采用在线检测和物料平衡已经实现流加。但在我国由于缺少在线检测仪表和装备落后,尚未普遍实现流加<sup>[2-4]</sup>。

本文提出另一种途径:流加的关键问题是要知道流加量。如果菌体的流加速率正好等于糖耗速率,那么问题就得到解决。

本文的结果是:解决了洁霉素发酵平衡生产期中糖耗速率估计问题。在我国我们首次实现了洁霉素发酵流加补料的计算机系统。

### 糖耗速率预测模型

#### (一)糖耗速率

在FBFLP(Fed-Batch Lincomycin Fermentation Processes)的平衡生产期中,三项主要生化参数——菌体浓度 $X$ 、基质浓度 $S$ 和产物 $P$ ,可以简化为成一项参数——基质浓度 $S$ 来研究。这是由于:菌体浓度 $X$ 近似恒定;产物 $P$ 的比生产速率由菌体饥饿状态决定,而饥饿状态又取决于菌体对糖的摄入率。因此只需研究有关糖参数就可以了。目前在生产中只有二项糖的瞬时参数——残糖浓度和补糖量。它们不能直接反映菌体的摄糖量。现有的机理模型是关于残糖浓度的常微分方程,即在 $\Delta t$ 时间内,有下列关系:

(残糖浓度变化速率)

$$= (\text{补糖量}) - (\text{菌体摄糖量})$$

在实际生产中,残糖浓度是可以测量

本文于1992年1月20日收到。

国家七五重点科技攻关项目“75-53-01-13/02”。

的,补糖量也是可以测量的,唯有摄糖量是由菌体决定的,是难以确定的。摄糖量具有决定比生产速率的重要作用。为了便于研究菌体的摄糖速率,有必要采用一个既等价于糖浓度又直观的重要状态变量。由于菌体的摄糖速率等于基质中糖的消耗速率,于是自然引伸出糖耗速率变量。

定义:糖耗速率为每个采样周期内的糖耗量。其计算公式为:

(糖耗速率)

$$= \frac{(\text{采样周期} - \text{残糖初值}) - (\text{采样周期} + \text{残糖终值}) + \text{补糖量}}{(\text{采样周期})} \quad (2)$$

## (二)糖耗速率预测

糖耗速率预测能解决残糖浓度难以检测和糖耗速率规律难以描述的问题。为了进行预测,需要解决以下几个问题:(1)由于8h采样间隔致使数据量少,如何才能从较少的数据中获得更多的信息。(2)糖耗速率时间序列的特征是什么?如何对它估计?这个估计能否导出一步甚至M步预测?(3)预测模型的机理解释。

从信息角度看,频域分析比时域分析得到的信息要多些,因此模型化过程应基于频域分析。具体步骤如下:

1. 对各批发酵过程的糖耗速率时间序列进行总体统计特性研究,可以确认它们是平稳序列。

2. 频谱或功率谱可以反映时间序列的特征:每个FBLFP都有它自己的特征。故对每个FBLFP糖耗速率时间序列要分别建模。建模的途径是:用时间序列的前一部分观测值进行谱密度估计,并且用该估计充当整个时间序列的估计,由它作出预测模型。

3. 在数据量少的情况下,宜采用极大熵谱估计<sup>[6]</sup>。

由时间序列的实现:  $\{X_t | t=1, 2, \dots, n\}$  可以知道熵(ENTROPY)为:

$$H(X) = -\sum P(X) \log P(X) \quad (3)$$

其中  $P(X)$  是联合概率密度,  $X$  的熵率为:

$$h = \lim \frac{H(X)}{N} \quad (4)$$

熵率的直观意义是时间序列  $\{X_t\}$  中每一个  $X_t$  的出现带来的平均信息。

由极大熵谱估计得出模型是阶次为  $P$  的自回归模型  $AR(P)$ 。

4. 从另一个方面来看,以谱密度估计为依据,采用频域方法——ЯГЛОМ方法,可以得到均方意义下的最优一步预测公式。这个公式也是采取  $AR(P)$  形式的。

5.  $AR(P)$  模型的定阶问题:由于FBLFP的时间序列观测值是逐步来到的,所以有必要对时间序列逐次定阶,即每增加一个观测值就对时间序列定阶一次。采用FPE(Final Prediction Error)准则确定自回归模型的最优价,直至最优阶不再变化为止。

6. BURG算法是自回归参数的递推算法:用它可以估计出  $AR(P)$  模型的参数。最优阶之下的自回归模型才可作为整个时间序列的预测模型。该预测模型又是均方意义下的最优一步预测公式。

模型化流程见图1,模型化依据见表1。

7. 为了保证上述模型化过程的正确性,需要分三个方面加以说明。第一个方面是作者应用有关方法得到了实际过程时间序列是平稳序列的结果,后二个方面引用文献,从两条不同途径说明了预测模型的形式是自回归模型。这三方面的工作分述如下:

(1) 平衡生产期糖耗速率时间序列

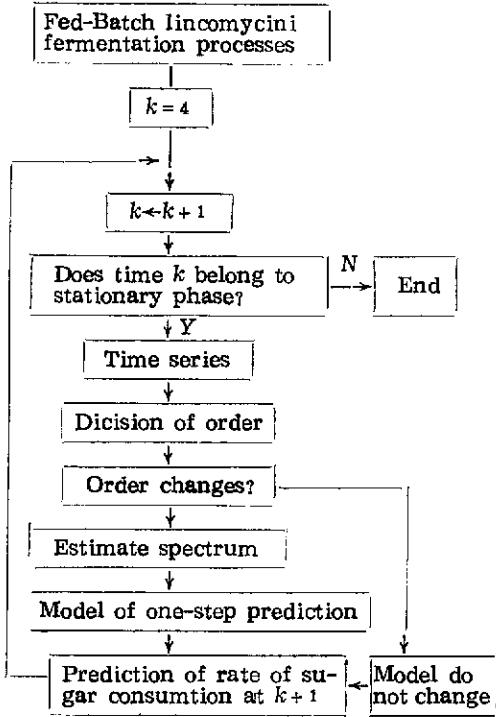


图 1 糖耗速率预测模型化流程图

Fig.1 Flowchart of prediction's modelling of rate of sugar consumption

表 1 模型化的依据

Table 1 Basis of modelling

Step	Goal	Principle	Method
1	Decision of order	Minimum variance of error one-step's prediction	Criterion of final prediction error
2	Estimate of spectrum	Maximum entropy	BURG's algorithm
3	One-step prediction	Optimal error of mean square	Яглом method

是平稳序列。众所周知，时间序列平稳性检验的方法如下<sup>[6]</sup>：

自相关函数为：

$$\rho_k = \frac{\text{Cov}(X_t, X_{t+k})}{\sigma_{x_t} \sigma_{x_{t+k}}} \quad (5)$$

由 $\{\hat{\rho}_k | k=1, \dots, n\}$ 可检查时间序列是否平稳序列，当 $k$ 值增大时，若 $\hat{\rho}_k$ 下降很快，则时间序列是平稳的，否则是非平稳的。

式(2—5)是理论的，实际应用时使用 $\rho_k$ 的估计——样本自相关函数：

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (6)$$

作者按上述方法对华北制药厂洁霉素发酵过程半年的数据进行了平稳性检验，得到了糖耗速率时间序列是平稳序列的结论。

(2) 在BURG极大熵准则下，预测模型是AR(P)的。

引理 1. 平稳序列 $\{X_t | t=1, 2, \dots, n\}$ ,  $t \geq p$ ,  $p$ 为由FPE准则确定的阶数，在极大熵准则下，谱密度估计为<sup>[5]</sup>：

$$\hat{f}_x(\lambda) = \frac{\theta_0^2}{2\pi |1 + \sum \varphi_k e^{-i k \lambda}|^2} = \frac{P_M}{|1 + \sum \varphi_k e^{-i k \lambda}|^2}$$

$\theta_0$ 是极大解析函数 $\frac{\Theta(Z)}{\varphi(Z)}$ 中 $\Theta(Z)$ 的

系数 $\theta_0 \in \{\theta_k | k=0, 1, 2, \dots, p\}$ ,  $P_M$ 为方差。

定理1. 极大熵准则下的预测模型是AR(P)的<sup>[5]</sup>。

(3) 在Яглом均方误差最优标准下，最优一步预测公式为AR(P)的。

引理2. (Яглом) 设 $X_t$ 是平稳序列，其谱密度为<sup>[7]</sup>：

$$f_x(\lambda) = f^*(e^{i\lambda})$$

$$f^*(Z)$$

$$= \left| \begin{array}{c} B_0 \\ A_0 \end{array} \right| \frac{(Z-b_1) \dots (Z-b_q)(1-b_1 \bar{Z}) \dots (1-b_q \bar{Z})}{(Z-a_1) \dots (Z-a_p)(1-a_1 \bar{Z}) \dots (1-a_p \bar{Z})}$$

其中  $0 < |a_i| < 1, 0 < |b_i| < 1$

当 $P \geq q$ 时，确定一个 $p-1$ 次的多项式 $Q(Z)$ ，使得

$$Q(Z) = Z^{p-1} (Z-b_1) \dots (Z-b_q)$$

以 $a_1, a_2, \dots, a_M$ 为根。

$$\text{令 } \Phi_M(Z) = \frac{Q(Z)}{Z^{M-1}(Z-b_1)\dots(Z-b_q)}$$

则已知 $\{X_t | S < 0\}$ 之下对 $X_M, M > = 0$ 的最优预测为:

$$\hat{X}_M = \int_{-\infty}^{\infty} \Phi_M(e^{\lambda}) dZ_x(\lambda)$$

并且记 $X$ 的预测误差为 $\sigma_M^2$ , 则

$$\lim \sigma^2 = \lim \|X_M - \hat{X}_M\|^2 = \int_{-\infty}^{\infty} f(\lambda) d\lambda = R_x(0)$$

定理2. 平稳序列 $X$ , 在均方误差标准下, 最优一步预测公式为AR(P) [7].

下面进行参数估计. 本文使用的BURG算法如下 [6]:

设过程为 $X_t; a_1, \dots, a_M$ 为自回归系数. 由Levison递推算法有:

$$\begin{pmatrix} 1 \\ a_1^{(M)} \\ a_2^{(M)} \\ \dots \\ a_M^{(M)} \end{pmatrix} = \begin{pmatrix} 1 \\ a_1^{(M-1)} \\ a_2^{(M-1)} \\ \dots \\ a_{M-1}^{(M-1)} \\ 0 \end{pmatrix} + C^{(M)} \begin{pmatrix} 0 \\ a_{M-1}^{(M-1)} \\ \dots \\ a_1^{(M-1)} \\ 1 \end{pmatrix}$$

其中 $a_i^{(j)}$ 上标表明系数 $a_i$ 属于AR(P)阶模型.

自回归系数的递推公式为:

$$\begin{cases} e_+^{(0)} = X_t, e_-^{(0)} = X_{t-1} (t = 1, 2, \dots, N) \\ C^{(M)} = \frac{-\sum a e_+^{(M-1)} e_-^{(M-1)}}{\sum [(e_+^{(M-1)})^2 + (e_-^{(M-1)})^2]} \\ e_+^{(M)} = e_+^{(M-1)} + C^{(M)} e_-^{(M-1)} \\ e_-^{(M)} = e_-^{(M-1)} + C^{(M)} e_+^{(M-1)} \\ a_i^{(M)} = a_i^{(M-1)} + C^{(M)} a_{M-i}^{(M-1)} \quad (i = 1, \dots, M-1) \\ a_M^{(M)} = C^{(M)} \end{cases}$$

模型的残差方差 $P_M$ 和自相关函数 $\hat{R}_x(M)$ 的递推式为:

$$\begin{cases} P_0 = \hat{R}_x(0) = \frac{1}{N} \sum_{t=1}^N X_t^2 \\ P_M = P_{M-1} (1 - C^{(M)^2}) \\ \hat{R}_x(M) = -C^{(M)} P_{M-1} - \sum_{i=1}^{M-1} a_i^{(M-1)} \hat{R}_x(M-i) \end{cases}$$

上述建模方法应用于工业试验, 情况如下.

### 材 料 和 方 法

#### (一) 菌种

*Streptomyces lincolnensis* var. 83-5, 由华北制药厂提供.

#### (二) 培养基组成(g/L):

淀粉20; 葡萄糖40; 豆饼粉25; 氯化

表 2 洁霉素百吨发酵罐试验数据

Table 2 Experimental data of fermentor in 100m<sup>3</sup>

Order of sample	4	5	6	7	8	9	10	11	12	13	14
Observed value*	-0.05	-0.25	0.18	-0.24	0.12	-0.05	-0.1	0.13	-0.2	-0.4	0.22
Predicted value*	-0.6	-0.2	0.27	-0.5	0.14	0	0	0.03	-0.11	0.13	0.1

\*except average 0.61

钠 5; 碳酸钠 8; 硫酸铵 2; 磷酸二氢钾 25; 硝酸钠 8; 硝酸胺 2. 此外每罐加玉米浆300L, 玉米油130L, 消沫剂20L.

#### (三) 培养条件

发酵罐内温度 30°C, 空气流量 4000

m<sup>3</sup>/h, 搅拌速度 60rpm, 培养基残糖浓度12g/L; 残氮浓度45g/L, pH6.8.

#### (四) 试验设备

100m<sup>3</sup>发酵罐两只, 和计算机流加控制系统.

## 结果和讨论

糖耗速率预测模型可用于平衡生产期开始第40h起到140h左右一段区间。(第40h前不适用是因为受BURG算法的限制)。实际预测情况较好。表2给出工业试验实例。利用MEBUGR软件, 通过计算机运算, 得出一步预测模型为:

$$\hat{X}_0 = -0.887X_{-1} - 0.433X_{-2} \quad (7)$$

在两个100m<sup>3</sup>罐中, 首次实现了洁霉素发酵过程计算机流加补料系统。投入生产后, 发酵单位、发酵指数和小时增长率可分别提高11.3%、8.64%和5.84%。在我国洁霉素发酵工业生产中常常有这样的现象: 在本次采样时刻, 过程的输入相同, 但在下次采样时刻, 过程的输出——糖耗速率却不同。因此可能认为糖耗速率具有不确定性。事实上, 从糖耗速率时间

序列来看, 它实际上不是一阶模型。糖耗速率时间序列是至少二阶的自回归模型, 无法用一阶规律来描述。故从生化机理上对模型的解释是: 由于高浓度基质和细胞自组织作用,  $\Delta K$  时间区间内菌体的摄糖量取决于  $\Delta K-1$  时间区间内细胞的饥饿程度, 而后者又取决于  $\Delta K-2$  时间区间内的摄糖量。因此糖耗速率自回归模型至少是二阶的。如果遇到基质浓度上下波动, 就会出现阶数更高的自回归模型。

### 符号说明

$\beta$	非生长关联系数(h <sup>-1</sup> )
$F$	补料流量(m <sup>3</sup> h <sup>-1</sup> )
$r_p$	产物形成率(kgm <sup>-3</sup> h <sup>-1</sup> )
$r_s$	基质糖耗速率(kgm <sup>-3</sup> h <sup>-1</sup> )
$r_x$	细胞增殖率(kgm <sup>-3</sup> h <sup>-1</sup> )
$S$	糖基质浓度(kgm <sup>-3</sup> )
$S_i$	流加液中的糖浓度(kgm <sup>-3</sup> )
$V$	发酵体积(m <sup>3</sup> )
$X$	活细胞浓度(kgm <sup>-3</sup> )
$\mu$	比生长速率(h <sup>-1</sup> )
$\mu_{max}$	最大比生长速率(h <sup>-1</sup> )

### 参 考 文 献

- [1] 武汉大学复旦大学生物系微生物教研室: 微生物学, 高等教育出版社, 北京, pp. 235—238, 1987.
- [2] Meneil, H.: Fermentation, Oxford University Press, Oxford, pp. 65—112, 1990.
- [3] 上海第三制药厂, 上海第四制药厂: 抗菌素生产, 化工出版社, 北京, pp. 93—95, 134—135, 1979年.
- [4] Cooney, Biotechnology and Bioengineering, Vol. XXV, pp. 257—269, 1983.
- [5] 杨位钦, 顾 岚: 时间序列分析与动态数据建模, 第二版, 北京理工大学出版社, 北京, pp. 226—250, 529—532, 1990.
- [6] 李卓立: 实用经济计量与经济预测, 清华大学出版社, 北京, pp. 25—27, 1981年.
- [7] 谢衷洁: 时间序列分析, 北京大学出版社, 北京, pp. 64—188, 1990.

## Prediction of Rate of Sugar Consumption of Stationary Phase of Lincomycin Fermentation

Xu Chuanyu<sup>1</sup> Fang Chongzhi<sup>1</sup> Ma Zhenhua<sup>2</sup>

(Departments of Automation<sup>1</sup> and Applied Mathematics<sup>2</sup>, Tsinghua University, Beijing 100084)

This paper suggests an efficient prediction model of the sugar consumption in the stationary phase of large scale FBLFP(Fed-Batch Lincomyciny Fermentation Processes), by which the sugar absorbing rate of the cell can be predicted. In FBLFP of our country the computer system for feeding is desinged and made. The product of lincomycin is increased by 5%.

**Key words** Fed-batch lincomycin fermentation processes; prediction of rate of sugar consumption; continuous feeding