第24卷 第2期
2010年3月

湖南工业大学学报
Journal of Hunan University of Technology

Vol.24 No.2
Mar. 2010

最优培养基配比方案的数学建模与求解

潘劲松

（湖南机电职业技术学院，湖南 长沙 410151）

摘　要：通过对实验数据的分析，用拟合和逐步回归的方法建立了最优配比方案的数学模型并进行了求解，所得结果与实验数据基本吻合，具有合理性。

关键词：最优配比；matlab；单因素准则；逐步回归

中图分类号：O242.1；TQ929+.1　　**文献标志码**：A　　**文章编号**：1673-9833(2010)02-0022-05

Mathematical Model Building and Derivation for Optimal Proportioning Design of Culture Medium

Pan Jingsong

（Hunan Mechanical & Electrical Polytechnic，Changsha 410151，China）

Abstract：An optimum ratio mathematical model is obtained through analysis of experimental data and by using the fitting and stepwise regression methods. The solution is basically consistent with the experimental data which proves reasonably.

Keywords：optimum ratio; matlab; single-element principle; stepwise regression

1　提出问题

表1为某发酵实验中培养基的碳源（C_1,C_2,C_3）和氮源（N_1,N_2,N_3,N_4）含量及IFN-γ（γ-干扰素）产量的实验数据。本文以表1的实验数据为基础，建立最优配比方案数学模型，通过选择碳源和氮源的种类及含量使IFN-γ产量达到最大值。

2　模型假设及说明

模型假设：

1）每个实验都是在相同的外界条件下进行的，且外界条件的变化对实验不会产生影响；

2）每个实验都是独立进行、互不影响的；

3）每个实验都是在所需的条件下顺利进行的；

4）每个实验培养基中有相同且数量不变的真菌；

5）每个实验培养基中的碳源、氮源和真菌都是均匀分布，且充满整个培养基；

6）每个实验中的真菌利用氮源和碳源生成IFN-γ的能力是一样的；

7）每个实验中氮源、碳源的含量同时成比例变化不影响氮源、碳源的转化率；

8）每个实验中不同碳源（氮源）之间都是独立的，不互相影响；

9）每个实验用到的数据都用同一的单位；

10）因为IFN-γ是单一类型的干扰素分子，在下面的讨论中考虑不同的碳源和氮源被吸收后都只生成一种相同类型的干扰素分子；

11）每个实验的培养基本身都含有一定数量的碳和氮；

12）在求解最优配比方案时，不同碳源（氮源）价格是相同的，或者说他们的价格差别对于问题而言是可以忽略的。

假设的说明：

对于假设1）~9），是为了保证实验具有可行性、

收稿日期：2009-09-15

通信作者：潘劲松（1968-），男，湖南津市人，湖南机电职业技术学院副教授，硕士，主要从事高等数学教学及课程改革，高职教育管理等工作，E-mail：pjs196855@126.com

可比性，在现实的情况下这些假设是可以做到的，符合实际的。

对于假设7），是为了便于从实验数据中了解、得到更多的信息，从而建立模型。在实际情况下，当2个反应物在其他条件不变时，浓度成比例改变，它们的转化率一般是不变的，在这里也假设是不变的。

对假设10），是为了在判断碳源和氮源的优劣进而选择合适的原料时，不需要去考虑因生成的IFN-γ有不同种类，而导致在判断碳源和氮源时，无法根据表中的量去判断的情况，在现实中因为IFN-γ是单一类型的干扰素分子，考虑其只会产生一种IFN-γ，也是合理的。

对于假设11），是从实际情况出发，考虑到一方面每个培养基中本身就含有必备的营养素，也就会含有一定的碳和氮；另一方面，因为实验中有5组数据是在只有碳或氮的情况下生成的，如果没有碳或氮是不会出现这些情况的。

对假设12），是为了在判断碳源和氮源的优劣进而选择合适的原料时，只需从相同原料时生成IFN-γ产量的多少考虑即可，这样假设是从方便建模的角度出发的。在现实中，会出现价格不同，甚至相差很大的情形，在这个时候就必须要考虑经济效益了，这将在模型改进中予以简要说明。

表1　培养基的实验数据

Table 1　Experiment data of medium

实验序号	碳源和氮源的质量/g							IFN-γ产量/g	实验序号	碳源和氮源的质量/g							IFN-γ产量/g
	C_1	C_2	C_3	N_1	N_2	N_3	N_4			C_1	C_2	C_3	N_1	N_2	N_3	N_4	
1	0.1	0	0	0	0.5	0	0	11.690 0	17	2	0	0	0	0	0	0	165.540 0
2	0.4	0	0	0	0.5	0	0	39.129 7	18	2	0	0	0.5	0	0	0	72.000 0
3	1	0	0	0	0.5	0	0	68.203 5	19	2	0	0	2	0	0	0	54.900 0
4	2	0	0	0	0.5	0	0	83.658 0	20	2	0	0	3	0	0	0	42.840 0
5	3	0	0	0	0.5	0	0	75.420 0	21	2	0	0	4	0	0	0	164.250 0
6	0	0.2	0	0	0.5	0	0	20.736 0	22	2	0	0	5	0	0	0	138.040 0
7	0	0.5	0	0	0.5	0	0	20.741 7	23	2	0	0	0	0	0	0	126.900 0
8	0	0.8	0	0	0.5	0	0	22.583 6	24	2	0	0	0	0.5	0.5	0	11.484 0
9	0	1.5	0	0	0.5	0	0	25.821 3	25	2	0	0	0	2	2	0	17.160 0
10	0	0	0.1	0	0.5	0	0	5.624 0	26	2	0	0	0	3	3	0	20.800 0
11	0	0	0.5	0	0.5	0	0	10.235 7	27	2	0	0	0	4	4	0	57.240 0
12	0	0	1.5	0	0.5	0	0	14.590 9	28	2	0	0	0	5	5	0	72.960 0
13	0	0	2.5	0	0.5	0	0	20.334 6	29	2	0	0	0	0	0	0.01	17.608 0
14	0	0	0	0	0.5	0	0	7.296 0	30	2	0	0	0	0	0	0.05	30.940 0
15	0	0	0	0	0.5	0	0	42.588 0	31	2	0	0	0	0	0	0.1	11.475 0
16	0	0	0	0	0.5	0	0	51.600 0	32	2	0	0	0	0	0	0.3	12.084 0

3　建立模型及求解

经过分析认为第14～17和23共5次实验为特殊情形，在方案1中暂不考虑，将在方案2的第2步中予以说明。

方案1　首先，在N_2含量不变的准则下，研究不同C_i（i=1,2,3）含量对产生IFN-γ的影响。主要思想如下：找出含C_i（i=1,2,3）源的项及其对应的IFN-γ产量项，利用数学软件matlab把表中的C_i（i=1,2,3）含量与IFN-γ产量有关的数据拟合成曲线，建立两者之间的函数关系[1]。将已建立的反映C_i（i=1,2,3）含量与IFN-γ产量关系的3条曲线放在同一坐标系中，进行观察、比较，在一定区间内，取最上方的曲线所对应的碳源为在该区间内所选择的碳源，对应区间内取得函数最大值时的自变量即为碳源的含量。

其次，在C_1含量不变的准则下，研究不同的N_j（j=1,2,3,4）含量对IFN-γ产量的影响，方法同上。

最后，根据在N_2含量不变的准则下求出的最佳碳源C_i（i=1,2,3），将N_2与C_i（i=1,2,3）的组合作为一种优化的配比方案。同理，根据在C_1含量不变的准则下求出的最佳氮源N_j（j=1,2,3,4），将C_1与N_j（j=1,2,3,4）的组合也作为一种优化的配比方案。

具体的做法如下：

实验设置为7组，实验1~5为第1组，实验6~9为第2组，实验10~13为第3组，实验15~18为第4组，实验1~5中把C_1按假设7）都化为2倍后所得新结果为第5组，实验23~28为第6组，实验29~32为第7组。

在N_2质量为0.5 g，N_1,N_3,N_4,C_2,C_3质量均为0时，用matlab对1~5组实验中C_1的含量x_1与IFN-γ产量y_1的关系进行3次曲线拟合，得拟合曲线如图1。

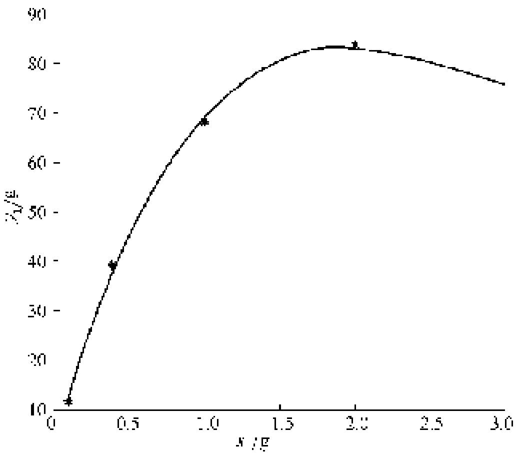

图 1　C_1 质量与 IFN–γ产量的拟合曲线

Fig. 1　The fitting curve of C_1 and IFN–γ

由拟合曲线推得拟合函数为：

y_1=5.168 3x_1^3−41.886 6x_1^2+103.462 5x_1+2.547 0。

同理，对固定的 N_2，得 C_2 的质量 x_2 与 IFN–γ产量 y_2 的拟合曲线如图 2，其拟合函数为：

y_2=−9.011 9x_2^3+23.719 0x_2^2−13.069 6x_2+22.473 3。

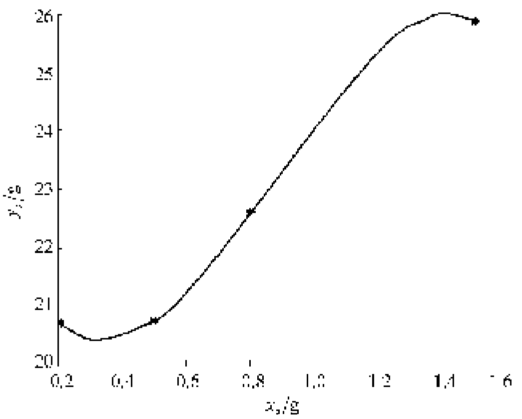

图 2　C_2 质量与 IFN–γ产量的拟合曲线

Fig. 2　The fitting curve of C_2 and IFN–γ

同理，对固定的 N_2，得 C_3 的质量 x_3 与 IFN–γ产量 y_3 的拟合曲线如图 3，其拟合函数为：

y_3=2.424 4x_3^3−10.215 6x_3^2+16.907 0x_3+4.033 0。

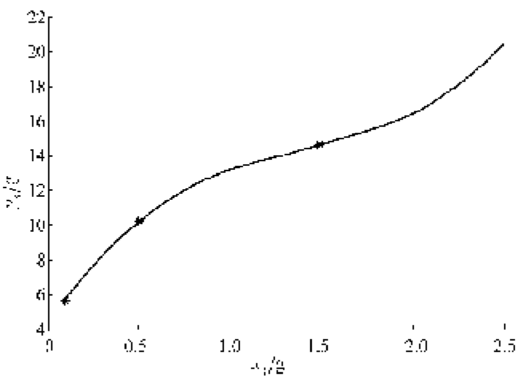

图 3　C_3 质量与 IFN–γ产量的拟合曲线

Fig. 3　The fitting curve of C_3 and IFN–γ

把上面建立的 C_1~C_3 质量与 IFN–γ产量关系的 3 条曲线图像拟合在同一坐标下，见图 4。

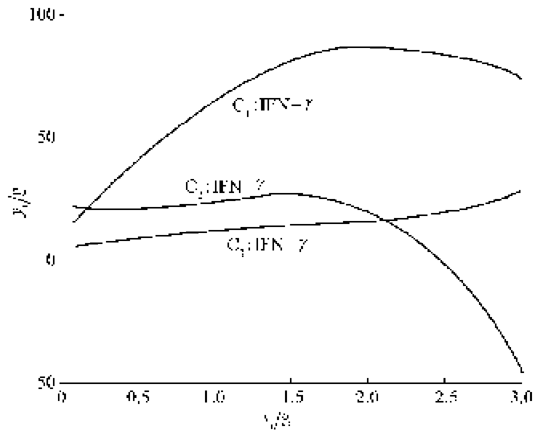

图 4　C_i 质量与 IFN–γ产量的拟合曲线

Fig. 4　The fitting curve of C_i and IFN–γ

从图 4 可知，以 N_2 质量不变为准则时，取 C_1 作为碳源，IFN–γ产量明显最大。因此，取 C_1,N_2 作为碳源和氮源。计算可得 C_1 最大值，该值即为 N_2 不变准则下 IFN–γ产量取最大值时 C_1 的质量，由此即得到一种优化配比方案，该方案为 m_{C_1}∶m_{N_2}=1.9∶0.5。

方案 2　方案 1 中只考虑了单一准则下各因素的最优条件，而实际上各个因素都在变化，要找到最优方案必须考虑周全。因此，应同时考虑 3 种碳源和 4 种氮源，分析它们对 IFN–γ产量的影响，从中选取影响较大且合理的组合。本文采用逐步回归分析法，建立逐步回归模型[2]，并用此方法选出了对 IFN–γ产量影响较大的碳源和氮源。具体操作过程如下。

第一步：分别将表 1 中 C_1~C_3,N_1~N_4 对应的 7 列数据记为列向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_7$，记 IFN–γ列的数据为列向量 $\boldsymbol{y}$。

令 $\boldsymbol{x}=[\,\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_7]$，调用 stepwise($\boldsymbol{x}$, $\boldsymbol{y}$)，即可得到第一次逐步回归诊断结果如图 5 所示。此图为 matlab 运算后输出结果的直接截屏。可以看出结果不理想，此时 p=0.022 9，取显著水平α=0.05 时，此回归模型可用，若取显著水平α=0.01 时，此模型不能用；R-square=0.461，较小；F=2.932，也较小。可见此回归模型不大合理，需要改进。

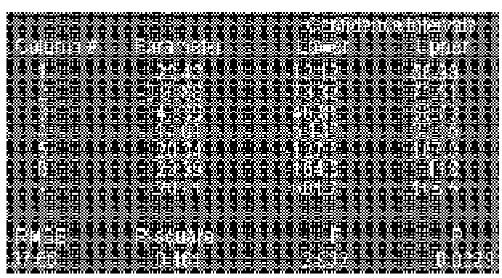

图 5　第一次逐步回归诊断结果

Fig. 5　Results 1 of stepwise regression diagnosis

第二步：仔细分析实验数据，发现 $\boldsymbol{x}$（14,:；16,:）中，每行向量相等，而 $\boldsymbol{y}$ 的值不等，说明此实验数据不可靠，且在现实中如果只用单一氮源，那么培养基将因缺乏碳源而崩溃，$\boldsymbol{x}$（17,:）和 $\boldsymbol{x}$（23,:）也是如此，故将其删除掉，用其余数据组合成新的 $\boldsymbol{x},\boldsymbol{y}$。重做第一步的操作，得第二次逐步回归诊断结果见图 6。

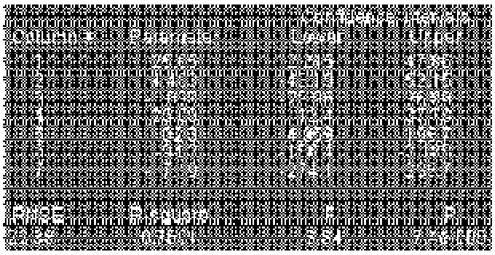

图 6　第二次逐步回归诊断结果

Fig. 6　Results 2 of stepwise regression diagnosis

可知此方案比第一步所得方案要好。p=7.4e−005，很小；R-square =0.765 1，比较大；F=8.84，也比较大。可见此方案可取。但 $\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_7$ 不显著，则应移去这 3 个向量，进一步诊断结果见图 7。

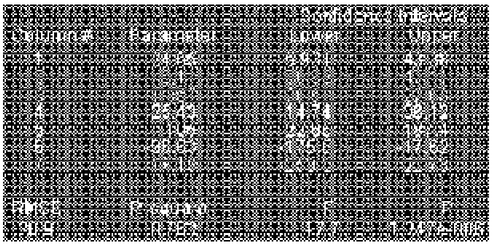

图 7　第三次逐步回归诊断结果

Fig. 7　Results 3 of stepwise regression diagnosis

由matlab易知：in =[1　4　5　6]; out =[2　3　7]。

从新的统计结果可以看出，虽然剩余标准方差 s（RMSE）没有太大的变化，但是统计量 F 的值明显增大，因此新的回归模型更好一些。

再运用matlab[3]知：

$\boldsymbol{x}$=[ones(27,1) $\boldsymbol{x}_1$' $\boldsymbol{x}_4$' $\boldsymbol{x}_5$' $\boldsymbol{x}_6$'];

[$\boldsymbol{b}$, bint, r, rint, stats]=regress($\boldsymbol{y}$, $\boldsymbol{x}$);

$\boldsymbol{b}$=[−32.617 1, 24.653 5, 25.428 9, 105.006 7, −96.634　0]^T;

stats=[0.763 0　17.702 6　0.000 0]。

至此可求出最优配比方案模型为

$$y = -32.617\,1 + 24.653\,5x_1 + 25.428\,9x_4 + 105.006\,7x_5 - 96.634x_6。$$

第三步：由试验数据可知，C_1,N_3 和 N_2 共存时，IFN−γ 产量比只有 C_1,N_2 时要低得多，可见最优方案中应该去掉 N_3，故将其删除掉，用其余数据组合成新的 $\boldsymbol{x},\boldsymbol{y}$，再次调用 stepwise($\boldsymbol{x},\boldsymbol{y}$)，得到逐步回归诊断结果见图 8。

由此时结果可知最优配比方案得到进一步改善，此时统计参数 p=2.996e−004，很小；R−square=0.780 8，比较大；F=8.903，也比较大。可见此方案可取。

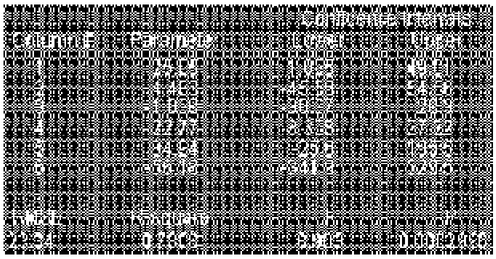

图 8　第四次逐步回归诊断结果

Fig. 8　Results 4 of stepwise regression diagnosis

但图 8 所示仍有些参数不显著，再重复第二步中的操作，最终得到较好的回归诊断结果，如图 9 所示。

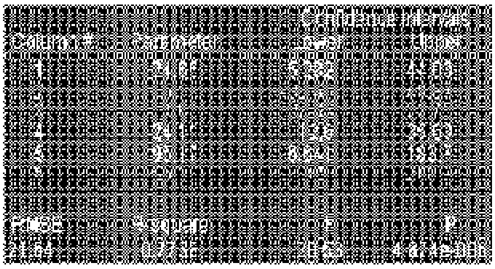

图 9　第五次逐步回归诊断结果

Fig. 9　Results 5 of stepwise regression diagnosis

从新的逐步回归诊断结果可看出，统计量 F 的值明显增大，因此这种回归模型比第二步所得更好。

再用 matlab 知：

$\boldsymbol{x}$=[ones(22,1) $\boldsymbol{x}_1$' $\boldsymbol{x}_4$' $\boldsymbol{x}_5$'];

[$\boldsymbol{b}$,bint,r,rint,stats]=regress($\boldsymbol{y}$, $\boldsymbol{x}$);

$\boldsymbol{b}$ =[−27.697 6, 24.653 5, 24.114 0, 95.167 8]^T;

stats =[0.773 8　20.526 0　0.000 0]。

至此求出最终的最优配比模型为：

$$y = -27.697\,6 + 24.653\,5x_1 + 24.114\,0x_4 + 95.167\,8x_5。$$

综上所述，碳源 C_1 和氮源 N_1,N_2 是影响 IFN−γ 产量的主要因素。所以原问题中要求得最优培养基配比方案，可以转化为在选取碳源 C_1 和氮源 N_1,N_2 作为原料后，确定它们的含量使得 IFN−γ 产量最大。

方案 3　通过上述讨论，已经选出了最优配比方案所需碳源和氮源的类型，下面具体讨论它们的含量问题。鉴于 N_2,N_3 共存时，IFN−γ 的产量低于仅有 N_2 时的情况，且不能断定 N_1,N_2 共存时 IFN−γ 的产量会达到最大，同时也没有 C_1,N_1,N_2 共存时的数据，因此通过分析，可以先任意给出 C_1 的含量，在 N_1,N_2 中选择一种，使之与 C_1 共存时，让 IFN−γ 的产量达到最大，从而求出氮的含量。

4　模型分析

对方案 1，主要是在确定 1 个因素的含量之后，研究其它单一因素的改变对目标的影响。运用这种方法，可粗略判断不同因素对目标的贡献大小，得到粗略的优化配比方案。这个模型的优点在于可直接利用

实验数据表中的数据进行比较而得到优化配比方案，缺点在于没有把实验数据表中的每个因素都进行比较，得到的结果只是局部最优，不一定是整体最优，同时在比较时没有考虑第14~17和23这5次实验，也会对所得结果造成一定影响。

对方案2，主要是用逐步回归分析法找出最优的碳源和氮源种类。运用这种方法得到的最优组合是比较合理的，它是根据实验数据考虑了全部变量的影响而得到的。

对方案3，是对方案2进行了补充说明。它在通过计算碳源、氮源的含量来获得最优方案时，参照了模型1的方法。这个模型的优点是解决了在取得回归方程之后，无法由实验数据表求出3个因素都在变动时的最优配比方案的难题，该模型通过确定1个因素之后，就可利用实验数据表求解最优配比方案了；缺点在于没有给出全部因素都在变动时的最优解。

在整个建模过程中，存在一个无法克服的困难：即模型是建立在实验数据之上的，由于实验数据具有随机性，这就决定了模型所得结果也具有随机性。

5 模型的改进

1）模型是建立在实验数据基础上的，运用的是数理统计的方法，可从增加实验测试次数来提高模型的准确度。

2）在实验设计时考虑运用正交实验设计。

3）模型假设12）不成立时，就要考虑到价格对选取碳源、氮源种类的影响。这种情况下，笔者建议在建立模型时把价格乘上质量来代替原模型中的质量，这样就可用上述模型去选择碳源、氮源种类及含量。

4）在建立模型时没有分析氮过量时对模型的影响，事实上如果氮源过量，会导致ph值升高，从而会破坏培养基，导致IFN-γ产量下降。因此建议在添加原料时控制好氮的含量。

5）在最优配比方案的碳源和氮源含量求解问题中，方案3给出了较特殊的解法，事实上，若可找到关于IFN-γ产量与碳源和氮源含量的函数关系，就可直接作为回归方程的约束条件来求解最值。

6）根据表1中的实验数据，由方案3求解，就可以通过二次回归得出最优配比方案如下：

a）当选择C_1,N_1时，其最优配比模型为：

$$y=0.736\,3+34.079\,7x_1+38.039\,7x_4-5.933\,5x_1^2-3.968\,9x_4^2,$$

可求得配比模型最优解为：

$x_1=4.151$，$x_4=2.604$，$y_{max}=113.281\,5$。

b）当选择C_1,N_2时，其最优配比模型为：

$$y=0.736\,3+34.079\,7x_1+38.039\,7x_5-5.933\,5x_1^2-3.968\,9x_5^2,$$

此配比模型最优解为：

$x_1=2.872$，$x_5=4.792$，$y_{max}=140.156\,1$。

c）当选择C_1,N_1,N_2时，其最优配比模型为：

$$y=-5.315\,2+20.604\,0x_1+53.904\,7x_4+51.436\,6x_5-1.948\,3x_1^2-9.584\,6x_4^2-6.612\,6x_5^2,$$

此配比模型最优解为：

$x_1=5.287\,7$，$x_4=2.812\,1$，$x_5=3.889\,3$，$y_{max}=224.975\,7$。

参考文献：

[1] 李尚志. 数学建模竞赛教程[M]. 南京：江苏教育出版社，1996.
Li Shangzhi. A Contest Course of Mathematical Modeling [M]. Nanjing：Jiangsu Education Publishing House，1996.

[2] 潘劲松. 大学数学应用基础[M]. 长沙：湖南教育出版社，2004.
Pan Jingsong. Application Basics of High Mathematics[M]. Changsha：Hunan Educational Publishing House，2004.

[3] 赵 静，但 琦. 数学建模与数学实验[M]. 北京：高等教育出版社，2002.
Zhao Jing，Dan Qi. Mathematical Modeling and Mathematical Test[M]. Beijing：Higher Education Press，2002.

[4] 潘劲松. “2+1”人才培养模式下高职数学课程改革的探讨[J]. 教育与职业，2008(10)：97-98.
Pan Jingsong. A Study on Higher Mathematics Course of '2+1' Personal Training Model[J]. Education and Vocational，2008(10)：97-98.

[5] 姜启源. 数学模型[M]. 北京：高等教育出版社，1987.
Jiang Qiyuan. Mathematical Modeling[M]. Beijing：Higher Education Press，1987.

[6] 雷功炎. 数学模型讲义[M]. 北京: 北京大学出版社，1999.
Lei Gongyan. A Lecture of Mathematical Modeling[M]. Beijing：Peking University Press，1999.

（责任编辑：李玉珍）