

# 相似准则方程的正确建立及其 与回归正交旋转设计的比较

蒋亦元

(东北农业大学)

**提 要** 该论文找出相似准则和相应的实验求出组分方程和组合成  $\pi$  关系式的预测误差较大的原因及其修正对策。指出 Murphy 的  $\pi$  函数合成理论中“组分方程必须具有相同型式”的约束和回归方程相关系数过低是误差产生的主要原因。在与回归正交旋转试验设计的对比中指出了  $\pi$  试验设计的一些特点, 其中回归试验结果时方程的模式几乎不受限止, 这一特点是保证预测精度的首要条件, 在此基础上提出了分区组合  $\pi$  方程和分区预测的新方法。

**关键词** 相似准则 组分方程  $\pi$  关系式 回归正交旋转设计

## Proper Formulation of $\pi$ - Equation and a Comparison With Regressive Orthogonal Rotating Experimental Design

Jiang Yiyuan

(Northeast Agricultural University, Harbin)

**Abstract** It is more often than not that prediction precision of an empirical equation is low, which is obtained through the  $\pi$ - equation, the component equations and the corresponding experiments. This paper revealed that “The component equations must have the same form”, advanced by G. Murphy and the lower correlation coefficient of regression model are the major sources of the prediction's error. Based on the experiment of “Settling Time of a Sphere in a Fluid” containing 7 variables, author explicitly suggested that the component equations may have different forms from each other and advanced that the prediction may be carried out in a set of separated areas on the curved surface in the factorial space. Each area has its own  $\pi$ - equation which is formulated by corresponding component equations. A comparison with the regressive orthogonal rotating experimental design was briefly made.

**Key words** Dimensionless product Component equation  $\pi$ - equation Regressive orthogonal rotating experimental design

### 1 引 言

通过相似准则并进行试验求出  $\pi$  关系式和经验公式, 这一方法常因误差过大而日益受到冷遇。经检验得知只要严格遵循关于“保持其他自变量的选定值不变和各个自变量都有变

收稿日期: 1996-04-25

蒋亦元, 教授, 哈尔滨市香坊区 东北农业大学工程学院, 150030

化的机会'两原则以及只要以高相关系数回归试验点,则在各组分方程上以及在其两侧附近进行预测,其精度是很高的。当偏离组分方程时,误差会增大,其幅度由曲线走向与曲率而定<sup>[1]</sup>。本文将分析研究误差产生的原因,提出在因子空间内曲面周缘部分误差较大时的修正办法以及与回归正交旋转设计比较时它所具有的特点。

## 2 误差产生原因分析

文献[1]已提出检验误差要分在组分方程的曲线上和在曲线以外两个步骤进行。由以下几点来说明其产生的原因。

### 2.1 回归方程的相关系数不够高

在工程问题中通常以组分方程乘积组合较合适,如果单个组分方程的失拟程度较高,总 $\pi$ 关系式的误差因是上述误差的连乘而增加得更大。图1, 2, 3说明前文球体沉降实验中含有沉降时间 $t$ 和沉降距离 $s$ 两个物理量的因变项 $\pi$ 随自变项 $\pi_E, \pi_S, \pi_d$ 变化时各试验点和以高相关系数拟合的方程和曲线1, 3, 5以及稍差的相关系数拟合的方程和曲线2, 4, 6。

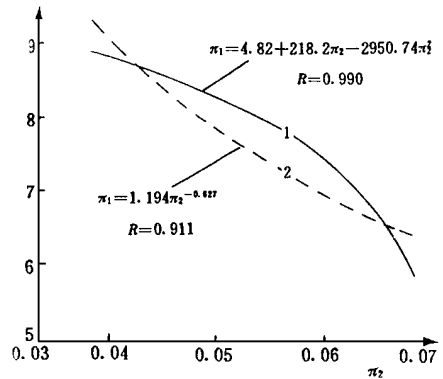


图1 由 $(\pi_1/2), \pi_2$ 的试验点回归的方程及曲线

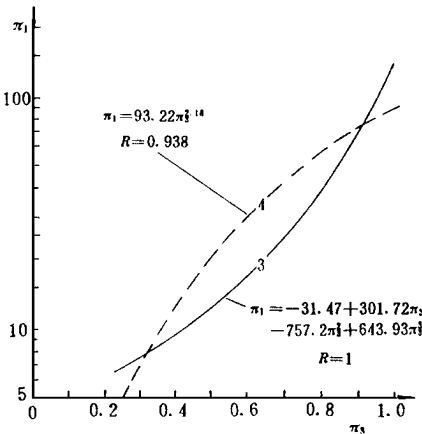


图2 由 $(\pi_1/3), \pi_s$ 的试验点回归的方程及曲线

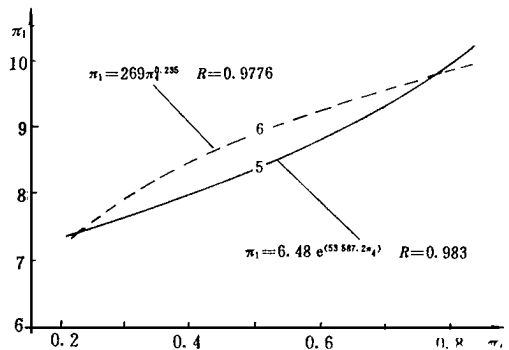


图3 由 $(\pi_1/4), \pi_d$ 的试验点回归的方程及曲线

当以方程2, 4, 6以乘积组合时其 $\pi$ 关系式

$$\pi = \frac{1.194 \pi_E^{0.627} \times 93.22 \pi_S^{2.18} \times 269 \pi_d^{0.275}}{8.91^2} \quad (1)$$

将物理量代入各 $\pi$ 项得经验公式

$$S = 0.0548 t^{1.647} d^{0.516} \rho_f^{1.795} \rho_s^{2.02} \mu^{-0.226} \quad (2)$$

各试验点的数值代入上式的  $S$  预测值, 误差结果见表 1。  $S$  预测值与实测值  $1m$  之差即为表中的相差率<sub>2</sub>, 而相差率<sub>1</sub> 即为文献[1]中表 5 中的值, 可见这两个相差率的差别竟达 40 ~ 50 倍之多, 一般的也有 20~ 30 倍之多。象由 2, 4, 6 方程作出的预测显然失去其指导意义。但从其相关系数看似似乎还不算太低, 所以可得出结论: 回归组分方程时, 必须具有很高的相关系数是最基本的要求。

表 1 组分方程的相关系数对总  $\pi$  关系式的预测误差的影响

$(\pi_{1/2})_{3,4}$			$(\pi_{1/3})_{2,4}$			$(\pi_{1/4})_{2,3}$		
$\pi_2$	相差率 <sub>1</sub> /%	相差率 <sub>2</sub> /%	$\pi_3$	相差率 <sub>1</sub> %	相差率 <sub>2</sub> /%	$\pi_4 (\times 10^{-5})$	相差率 <sub>1</sub> /%	相差率 <sub>2</sub> /%
0.0378	0.20	25.0	0.989	0.35	2.20	0.830	1.07	21.2
0.0478	1.40	19.7	0.660	0.51	25.3	0.616	0.20	25.0
0.0578	0.34	17.9	0.420	0.51	46.3	0.499	0.76	27.3
0.0678	0.89	26.8	0.340	0.21	50.4	0.317	1.34	21.6
-	-	-	0.252	0.48	12.2	0.211	0.23	19.5
R	0.990	0.911	R	1	0.935	R	0.983	0.977

### 2.2 受到现有定理的不恰当约束而失拟

在按试验结果拟合各个组分方程时受到“两个组分方程必须具有相同的型式”<sup>[1]</sup>的约束, 使得各组分方程本应为不同型式时也得勉强去求同。这种勉强求同就是造成上例中相关系数更差, 预测误差更大的另一个原因。

较早提出相似准则函数理论的是 Glenn Murphy, 在列出两个组分方程由乘积合成总的  $\pi$  关系式时提出了与上述相同的条件  $F(\pi_2, \pi_3) = \frac{F(\pi_2, \pi_3) F(\pi_2, \pi_3)}{F(\pi_2, \pi_3)}$  indicates that the two component equations must have the same form<sup>[3]</sup>。

而在文献[1]的球体沉降的实例中就并未受此约束, 用不同型式的方程去拟合, 结果预测的精度很高, 即表 1 的相差率<sub>1</sub>。

### 2.3 可以用不同型式的组分方程进行组合的证明

假如有两个组分方程, 一个为幂函数  $\pi_{1/2} = a + b \pi_2^c$  此时  $\bar{\pi}_2 = g$ ; 另一个为指数函数  $\pi_{1/3} = d + e^f \pi_3$ , 此时  $\bar{\pi}_3 = h$ , 则乘积合成为

$$F(\pi_2, \pi_3) = \frac{(a + b \pi_2^c)(d + e^f \pi_3)}{K}$$

式中  $a, b, c, d, f$  —— 方程的回归系数和指数;  $g, h$  —— 做试验所选定的值。

当  $\bar{\pi}_3 = h$  时,  $F(\pi_2, \pi_3) = \frac{a d + a e^f \pi_3 + d b h^c + b h^c e^f \pi_3}{K}$ , 由试验和  $\pi$  函数理论得知此时的  $K$  值为  $\pi_3 = g$  代入  $\pi_{1/3}$  所得的值, 即  $K = d + e^f g$ 。同理当  $\bar{\pi}_2 = g$  时,  $K$  为  $\pi_2 = h$  代入  $\pi_{1/2}$  所得值, 即  $K = a + b h^c$ 。而且这两个值在试验时是在因子空间相同坐标点上的同一个因变量值, 所以这两个  $K$  是同值, 则

$$F(\pi_2, \pi_3) = \frac{(d + e^f \pi_3)(a + b h^c)}{d + e^f g} = d + e^f \pi_3$$

这说明当  $\pi_h = h$  代入总  $\pi$  关系式时, 即进行“降维”分析时,  $\pi_{h/3}$  可以获得原来的型式, 这证明以不同型式的方程进行组合是可行的。

允许不同型式的方程进行组合, 就具有重大意义, 因为它扩大了组合方程—— $\pi$  关系式对自然界千姿百态的变化规律的反映能力。

### 3 相似准则函数理论的试验设计 (简称 $\pi$ 试验设计) 与二次回归正交旋转组合设计 (简称 ROR 设计) 的比较

ROR 设计在农业工程的科研和技术工作中是广泛应用的试验设计。由于具有“均匀分散性”、“整齐可比性”; 又具有“预测值的方差不依赖于试验点在因子空间中的位置”, ROR 试验设计既能“比较全面地反映各因素水平对指标影响的大致情况”又能“使各因素在不同水平时所对应的预测值具有相同的方差从而可直接比较”等优点<sup>[4]</sup>, 再加对所得方程以及回归系数尚可进行统计分析等, 这些优点都是以古典回归为基础的  $\pi$  试验设计所不及的。但  $\pi$  试验设计具有以下特点:

1) 试验次数少, 减少试验工作量和成本。如文献[1]球体沉降实验中有 6 个自变量, 3 个自变  $\pi$  项, 6 个组分方程, 若每个方程由 4 个试验点组成, 共需 24 次试验。而 ROR 设计, 当有 6 个变量时需作 100 次, 5 个变量时需 59 次试验。如果 1/2 实施, 分别也得作 59 和 36 次试验<sup>[4]</sup>。

2) 安排试验点不如 ROR 设计那样要求严格。如 ROR 设计, 在变量  $m = 3$ , 要求兼有正交性和旋转性时, 星号臂长  $r = 1.682$ 。如果没有达到, 而  $r = 1.352$  时, 就失去了旋转性, 仅有正交性。这两个  $r$  值仅差 20%。而在农业物料的试验中有些指标如湿度等是很难控制得那么精确的。

3) 二次回归 ROR 设计虽然由于正交性, 能比较全面地反映各因素水平对指标的影响, 但二次方程的型式本身总有其局限性。而  $\pi$  试验设计中的组分方程的型式在一般常用函数情况下可不拘泥于相同的型式, 对自然界的规律就有较广泛的描述能力, 而且当“降维”分析时又能还原到试验所得的方程。

如下式是用 Burgers 模型来描述水田土壤的流变方程<sup>[5]</sup>。土壤的应变

$$y = \tau \left[ \frac{1}{a} + \frac{1}{b} (1 - e^{-c t}) + \frac{t}{d} \right] \quad (3)$$

式中  $\tau$ —土壤应力;  $t$ —载荷作用时间;  $a, b, c, d$ —土壤特性常数。图 4 中的曲面表达了当  $t$  和  $\tau$  为变量时  $y$  的变化。

描述这样复杂的物理现象的方程(3), 很清楚是由两个方程组合的, 一个是  $y$  与  $\tau$  的线性方程, 另一个是  $y$  与  $t$  的指数方程, 如果用  $\pi$  试验设计通过试验求水田土壤的应变值的方程时, 从试验点的走向也能明显看出是由这两种不同型式

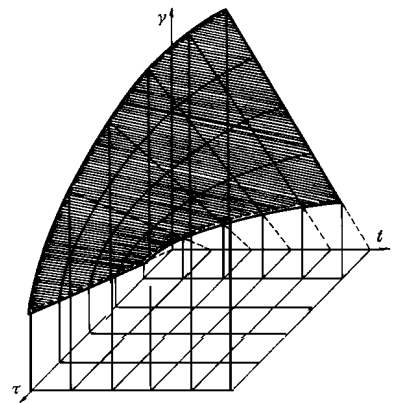


图 4 流变型土壤的应力—应变—时间图

的组分方程组合的。

### 4 分区组合 $\pi$ 方程并分区预测

现已明确: 组分方程的型式不受限制, 只要具有较高的相关系数, 组合成的  $\pi$  方程在组合方程的曲线上以及其周围具有很高与较高精度的预测功能。据此就可作如下的改进以提高整体的预测精度: 以图 5 所示的只有两个自变  $\pi$  项的问题为例来说明。

1) 把原先由 2 个改为由 4 个组分方程组合成  $\pi$  关系式。每个自变  $\pi$  项有 2 个选定值, 要在试验范围内分布适当, 通过试验获得 4 条曲线和 4 个组分方程, 即

$$(\pi_{1/2})_{\bar{3}}, (\pi_{1/2})_{\bar{3}}, (\pi_{1/3})_{\bar{2}}, (\pi_{1/3})_{\bar{2}}$$

由此可组合成 4 个  $\pi$  关系式

$$F_1(\pi_{\bar{2}}, \pi_{\bar{3}}) = \frac{(\pi_{1/2})_{\bar{3}}(\pi_{1/3})_{\bar{2}}}{F(\pi_{\bar{2}}, \pi_{\bar{3}})}, \quad F_2(\pi_{\bar{2}}, \pi_{\bar{3}}) = \frac{(\pi_{1/2})_{\bar{3}}(\pi_{1/3})_{\bar{2}}}{F(\pi_{\bar{2}}, \pi_{\bar{3}})},$$

$$F_3(\pi_{\bar{2}}, \pi_{\bar{3}}) = \frac{(\pi_{1/2})_{\bar{3}}(\pi_{1/3})_{\bar{2}}}{F(\pi_{\bar{2}}, \pi_{\bar{3}})}, \quad F_4(\pi_{\bar{2}}, \pi_{\bar{3}}) = \frac{(\pi_{1/2})_{\bar{3}}(\pi_{1/3})_{\bar{2}}}{F(\pi_{\bar{2}}, \pi_{\bar{3}})}$$

它们分别在以 4 个交点 , , , 为中心 的范围内进行较为准确的预测。此即所谓分区组合, 分区预测。

2) 由于它能较准确地预测, 就可省去对  $\pi$  关系式有效性的检验和原来用于检验的  $\pi_{\bar{2}}, \pi_{\bar{3}}$  的两组试验。因此用分区组合, 分区预测在只有两个自变  $\pi$  项时不会增加试验的工作量。而组合的  $\pi$  关系式增加 3 个, 计算工作量会有所增加, 但对用电子计算机计算而言是微不足道的。

### 5 结 论

在工程问题中常用乘积组合  $\pi$  关系式, 组分方程的相关系数必须要高, 组分方程可以具有不同的型式, 而且方程选型准确是获得高相关系数和预测精度的首要条件。分区组合  $\pi$  关系式进行分区预测可提高整体的预测精度。

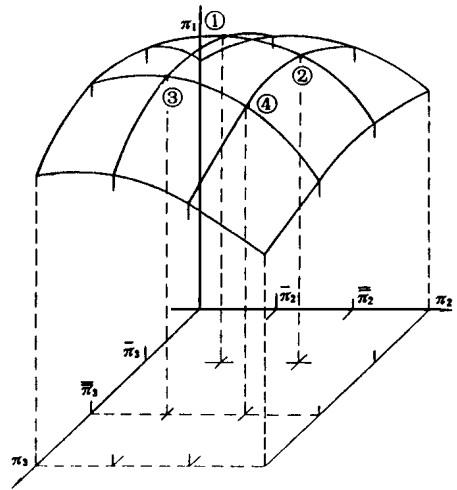


图 5 分区组合  $\pi$  关系式

### 参 考 文 献

- 1 蒋亦元, 涂澄海, 刘道顺等. 以球体的沉降实验检验  $\pi$  关系式预测的可行性. 农业工程学报, 1996, 12(3): 1~ 6
- 2 徐 挺 相似理论与模型试验. 北京: 中国农业机械出版社, 1982. 81, 86
- 3 Glenn Murphy. Similitude in Engineering. New York: The Ronald Press Company, 1950 302p
- 4 徐中儒. 农业试验最优回归设计. 哈尔滨: 黑龙江科学技术出版社, 1988. 68~ 133
- 5 潘君拯 流变学与水田土壤 自然杂志, 1982, 6(11): 817~ 822
- 6 李之光 相似与模化(理论及应用). 北京: 国防工业出版社, 1982 96~ 97