第44卷 第4期　　气 象 学 报　　Vol. 44, No. 4
1986 年 11 月　　ACTA METEOROLOGICA SINICA　　Nov., 1986

模糊集合嵌入岭回归在水稻产量年景气象预测中的应用*

李 耀 先

（广西壮族自治区气象台）

提　　要

本文根据模糊集合和岭回归理论，将模糊集合嵌入岭回归方程中，建立了玉林地区早稻产量年景的气象预测方程。

一、引　　言

目前，在我国开展的作物产量气象预测工作中，预测的模式有多种多样，但均以多元回归分析建立的模式较为广泛。近年来，模糊数学和岭回归分析发展较快，日益成为气象部门的数学工具。模糊数学在处理模糊性的事件中，其效果较令人满意；而岭回归则是在变量较多，且彼此间有很强的共线性时，选择的因子和确定的回归方程，其方差和系数的波动要较最小二乘法意义下的多元回归分析小[1,2]。本文根据模糊集合和岭回归的理论，将模糊集合嵌入岭回归方程中，从多元共线性的变量中，选择最佳的因子来建立玉林地区早稻产量年景的预测方程，其效果较理想。

二、因子的选择及隶属方程的确定

我们选用了玉林地区 1966 年至 1982 年共 17 年的早稻亩产量，经过调和权重处理后，求出其各年的相对产量(Y)。然后，用数理统计方法，选择了与相对产量关系比较密切的 10 个大气环流因子：头年 9 月西太平洋副高面积指数(x_1)；头年 9 月欧洲 W型环流出现天数 (x_2)；头年 10 月欧洲 W 型环流出现天数 (x_3)；头年 9 月欧洲 E 型环流出现天数 x_4)；3 月份亚洲纬向环流指数(x_5)；3 月份亚洲地区平均纬向环流指数(x_6)；3 月份 4 区 500 hPa 关键区距平(x_7)；头年 4 月 4 区 500 hPa 关键区距平(x_8)；头年 11 月份 9 区 500 hPa 关键区距平(x_9)；头年 12 月 4 区 500 hPa 关键区距平(x_{10})。

根据模糊集合原则[3]，建立各个预报因子(x_i)与相对产量(y)的隶属方程：

$$\tilde{\mu}_y(x_1)=\begin{cases}0, & x_1\leqslant 9\\ \dfrac{1}{1+e^{-0.6766(x_1-16)}}, & 9<x_1<22\\ 1, & x_1\geqslant 22\end{cases} \tag{1}$$

* 本文于 1985 年 4 月 29 日收到，1985 年 9 月 30 日收到修改稿。

$$\tilde{\mu}_y(x_2)=\begin{cases}0 & ,\ x_2\leqslant 3\\ 0.5245\ln x_2-0.6549 & ,\ 3<x_2<22\\ 1 & ,\ x_2\geqslant 22\end{cases}\tag{2}$$

$$\tilde{\mu}_y(x_3)=\begin{cases}1 & ,\ x_3\leqslant 5\\ 1.6316-0.1262\,x_3 & ,\ 13>x_3>5\\ 0 & ,\ x_3\geqslant 13\end{cases}\tag{3}$$

$$\tilde{\mu}_y(x_4)=\begin{cases}1 & ,\ x_4\leqslant 1\\ \dfrac{1}{1+e^{0.4499(x_4-9.5)}} & ,\ 1<x_4<17\\ 0 & ,\ x_4\geqslant 17\end{cases}\tag{4}$$

$$\tilde{\mu}_y(x_5)=\begin{cases}0 & ,\ 1.25\leqslant x_5\leqslant 1.67\\ 1.35-6.2448\ln x_5 & ,\ 1.25>x_5>1.06\\ 1 & ,\ x_5\leqslant 1.06, x_5\geqslant 1.68\end{cases}\tag{5}$$

$$\tilde{\mu}_y(x_6)=\begin{cases}0 & ,\ x_6\geqslant 1.58\\ 1-2.4871\ln x_6 & ,\ 1.58>x_6>1.03\\ 1 & ,\ x_6\leqslant 1.03\end{cases}\tag{6}$$

$$\tilde{\mu}_y(x_7)=\begin{cases}0 & ,\ x_7\leqslant -25\\ \dfrac{1}{1+e^{-0.0699(x_7-15)}} & ,\ -25<x_7<86\\ 1 & ,\ x_7\geqslant 86\end{cases}\tag{7}$$

$$\tilde{\mu}_y(x_8)=\begin{cases}1 & ,\ x_8\leqslant -65, x_8\geqslant 151\\ \dfrac{1}{1+e^{0.0432x_8}} & ,\ -65<x_8<30\\ 0 & ,\ 30\leqslant x_8<151\end{cases}\tag{8}$$

$$\tilde{\mu}_y(x_9)=\begin{cases}0 & ,\ x_9\geqslant 12\\ \dfrac{1}{1+e^{0.2106x_9}} & ,\ 12>x_9>-15\\ 1 & ,\ x_9\leqslant -15\end{cases}\tag{9}$$

$$\tilde{\mu}_y(x_{10})=\begin{cases}0 & ,\ x_{10}\leqslant 15\\ 0.8921\ln x_{10}-2.4940 & ,\ 15<x_{10}<48\\ 1 & ,\ x_{10}\geqslant 48\end{cases}\tag{10}$$

$\tilde{\mu}_y(x_i)$的意义为：$\tilde{\mu}_y(x_i)$值愈大，早稻产量就愈高；$\tilde{\mu}_y(x_i)$值愈小，早稻产量就愈低。

为了增加预报对象的信息，对预报对象(y)的定义域拓展为丰、平、欠三个产量年景等级，其隶属方程为：

$$\tilde{\mu}(y)=\begin{cases}0 & ,\ y\leqslant 97\%(\text{欠})\\ 0.1667\,y-16.1667 & ,\ 97\%<y<103\%(\text{平})\\ 1 & ,\ y\geqslant 103\%(\text{丰})\end{cases}\tag{11}$$

将各个因子的历年值分别代入上述相应的隶属方程，得出各年的 $\tilde{\mu}_y(x_i)$及 $\tilde{\mu}(y)$ 值（见表 1）。

表 1 $\tilde{\mu}_y(x_i)$及 $\tilde{\mu}(y)$值

项目 / 年代	$\tilde{\mu}_y(x_1)$	$\tilde{\mu}_y(x_2)$	$\tilde{\mu}_y(x_3)$	$\tilde{\mu}_y(x_4)$	$\tilde{\mu}_y(x_5)$	$\tilde{\mu}_y(x_6)$	$\tilde{\mu}_y(x_7)$	$\tilde{\mu}_y(x_8)$	$\tilde{\mu}_y(x_9)$	$\tilde{\mu}_y(x_{10})$	$\tilde{\mu}(y)$
1966	0.3599	0.6904	0.4958	0	0	0.2721	0.1238	0	0.1306	0.4125	0.2866
67	0.9091	0.6028	0.6220	0.7549	0.4231	0.2536	0.8127	0.7382	1	0.8614	1
68	0.0307	0	0	0	0	0	0	0	0	0	0
69	0.0533	0.0722	0	0.0777	0	0.3095	0	0.4892	0.2204	0	0
70	1	0.7655	0.4958	0.8284	1	1	1	1	1	0.6252	1
71	0.9467	0.4975	0.1172	0.4440	1	0.8551	0.8906	0.2298	0.8137	0.8819	0.7200
72	0.2403	0.9420	0	0.9669	0	0.5465	0	0.9245	0.1565	0.2220	0.3199
73	0.1510	0.2849	0	0.2451	1	0.9025	0.3167	0	0.9502	0.0845	0
74	1	0.6904	0	0.5560	1	1	1	0.8748	1	1	1
75	0.0909	0.4358	0.8744	0	0	0.5674	0.4306	0	0.1085	0	0.3699
76	0.5000	0.9420	1	0.5560	1	1	1	1	0.2204	0.2220	0.7700
77	0.0307	0.0722	1	0.0510	1	0	0.5865	1	1	0	0.6700
78	0.8490	0.1893	0	0.8834	0.5318	0.5465	0.9920	0.9275	0.8694	0.4787	0
79	0.8490	1	0.1172	1	0.6423	0.5054	0.9685	0.9080	0.0898	0.5695	0.6200
80	1	1	0.2434	0.9669	1	1	0.8762	0.7212	0.8137	0.8189	1
81	0.9467	0.4358	0.2434	0.5560	0	0.0588	0	0	0	0.9215	0
82	1	0.8895	0.6220	0.8834	0.0572	0.5883	0.0700	1	1	1	1

三、多元共线数据的分析

1. 多元共线性检测

用矩阵 $X=(x_{ij})_{m\times n}$ 和 $y=(y_1,y_2,\cdots,y_m)'$ 表示观测数据，$\tilde{X}=(\tilde{x}_{ij})_{m\times n}$ 和 $\tilde{y}=(\tilde{y}_1,\tilde{y}_2,\cdots,\tilde{y}_m)$ 表示标准化后的数据(其目的是消除自变量的量纲对因变量的影响，另一方面还可使设计构成一个正交子集)，其中：

$$\tilde{x}_{ij}=\frac{x_{ij}-\bar{x}_j}{\sigma_j}(i=1,2,\cdots,m;j=1,2,\cdots,n) \tag{12}$$

$$\tilde{y}_i=\frac{y_i-\bar{y}}{\sigma_y}(i=1,2,\cdots,m) \tag{13}$$

$\bar{x}_j$ 和 $\bar{y}$ 为样本均值，σ_j 和 σ_y 为标准差。经标准化处理后，$(\tilde{X}'\tilde{X})$ 便是自变量的相关系数矩阵 $R=(\tilde{X}'\tilde{X})=(\gamma_{ij})_{n\times n}$，其元素 γ_{ij} 为变量 x_i 与 x_j 的单相关系数，$(\tilde{X}'\tilde{Y})$便是自变量与因变量的相关系数矩阵 $W=(\tilde{X}'\tilde{Y})=(\gamma_{jy})_{n\times y}$，其元素 γ_{jy} 是自变量 x_j 与 y 的单相关系数。

设 P_j 为 $R=(\tilde{X}'\tilde{X})$的特征向量，λ_j 为对应的特征值，则有：

$$(\tilde{X}'\tilde{X})P_j=\lambda_jP_j \tag{14}$$

如果 $\lambda_j=0$，则 $(\tilde{X}'\tilde{X})P_j=0$，现以 $M(\tilde{X}'\tilde{X})$ 表示 $\tilde{X}'\tilde{X}$ 列向量张成的子空间，那末 $M(\tilde{X}'\tilde{X})=M(\tilde{X}')$。事实上，显然有 $M(\tilde{X}'\tilde{X})\subset M(\tilde{X}')$。另一方面，对任意向量 U，若 $U'\tilde{X}'\tilde{X}=0$，则必有 $U'\tilde{X}'\tilde{X}U=0$，即 $U'\tilde{X}'=0$。这表明与 $\tilde{X}'\tilde{X}$ 的列向量垂直的向量必与 $\tilde{X}'$ 的列向量垂直，故 $M(\tilde{X}')\subset M(\tilde{X}'\tilde{X})$。综合起来，即有 $M(\tilde{X}'\tilde{X})=M(\tilde{X}')$。根据这个事实，我们得到：

$$\tilde{X}P_j=0 \tag{15}$$

这就是说在自变量之间存在如下严格的线性关系：

$$\sum_{k=1}^{t} x_{kj} P_{kj} = 0 \qquad j = 1,2,\cdots,n$$

$$P_j = (P_{1j}, P_{2j}, \cdots, P_{tj}) \tag{16}$$

在实际问题中，若 $\lambda_j \approx 0$，则 $P_j \approx 0$，由(16)式知，自变量 $x_1, x_2, \cdots, x_n$ 有多元共线关洲：

$$\sum_{k=1}^{t} P_{kj} x_k \approx 0 \tag{17}$$

一般，从模拟试验中得出的看法是当 $\lambda \leqslant 0.05$ 时，可以放心地认为它是近似等于 0，而在 $\lambda \geqslant 0.10$ 时则否[4]。

本例由表 1 求出的$(\tilde{X}'\tilde{X})$、$(\tilde{X}'\tilde{Y})$相关系数矩阵见表 2 。

表 2　增广相关系数矩阵表

i \ j	1	2	3	4	5	6	7	8	9	10	y
1	1	0.5517	−0.1070	0.7465	0.3259	0.4205	0.5417	0.3775	0.4246	0.9448	0.5892
2		1	0.1514	0.6913	0.1789	0.5510	0.3198	0.4412	0.0121	0.5116	0.6384
3			1	−0.1906	0.0953	−0.0742	0.1538	0.2036	0.0820	−0.1267	0.4256
4				1	0.2325	0.4228	0.4293	0.6619	0.2902	0.6208	0.4646
5					1	0.6390	0.7963	0.3956	0.6308	0.1911	0.5111
6						1	0.5873	0.3033	0.4046	0.2938	0.4817
7							1	0.5150	0.4852	0.3396	0.5913
8								1	0.4469	0.2396	0.6068
9									1	0.4081	0.5571
10										1	0.5762

由表 2 求出$(\tilde{X}'\tilde{X})$的特征值为：$\lambda_1 = 4.6760$，$\lambda_2 = 1.7020$，$\lambda_3 = 1.1700$，$\lambda_4 = 0.9090$，$\lambda_5 = 0.7530$，$\lambda_6 = 0.4430$，$\lambda_7 = 0.1560$，$\lambda_8 = 0.1110$，$\lambda_9 = 0.0700$，$\lambda_{10} = 0.0090$。λ_{10}的特征值很小，近似为 0，且 λ_1 是 λ_{10} 的 519.6 倍，表明数据高度共线性。从 λ_{10} 的特征值所对应的特征向量发现，x_1, x_6, x_{10} 之间存在很强的共线性：

$$0.6910\ \tilde{x}_1 - 0.1190\ \tilde{x}_6 - 0.5980\ \tilde{x}_{10} \approx 0 \tag{17}$$

2．多元共线性的效应

通常的线性回归模型为：

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{ti} + \varepsilon_t (t = 1,2,\cdots,m) \tag{18}$$

引入标准回归系数 $\tilde{\beta} = (\beta_1, \beta_2, \cdots, \beta_n)'$，其中

$$\tilde{\beta}_i = \beta_i \sqrt{\frac{s_{yy}}{s_{ii}}}, \qquad (i = 1,2,\cdots,n) \tag{19}$$

这里，离差平方和 s_{ii} 和 s_{yy} 为：

$$s_{ii} = \sum_{t=1}^{m} (x_{ti} - \bar{x}_i)^2, \qquad s_{yy} = \sum_{t=1}^{m} (y_t - \bar{y})^2 \tag{20}$$

假定自变量相关矩阵 $R=(\tilde{X}'\tilde{X})$的特征值 $\lambda_1\geqslant\lambda_2\geqslant\lambda_3\geqslant\cdots\geqslant\lambda n$,相应的特征向量分别为 $P_1,P_2,P_3,\cdots,P_n$,则可证得回归系数 $\tilde{\beta}$ 的最小二乘估计 $\hat{\beta}=(\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}n)$为

$$\hat{\beta}=(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}=\sum_{k=1}^{n}\lambda^{-1}{}_kc_kP_k \tag{21}$$

其中
$$c_k=P_k'\tilde{X}'\tilde{y} \tag{22}$$

且有
$$E(\hat{\beta})=\tilde{\beta} \tag{23}$$

$$Var(\hat{\beta})=\sum_{k=1}^{n}\lambda^{-1}{}_k\sigma_y^2P_kP_k' \tag{24}$$

由(21)式可知，除非c_k接近于零，近似为零之特征值所产生的乘子 $\lambda^{-1}{}_k$ 将使多元共线变量相应回归系数的最小二乘估计取得很大值，而且回归系数具有与 P_k 相同的代数符号。或者说，高度多元共线性导致矩阵$(\tilde{X}'\tilde{X})$小的特征值，从而大的 $\lambda^{-1}{}_k$ 值使相应特征向量 P_k 在回归系数的最小二乘估计中占支配地位。由(24)式知，多元共线性也通过乘子 $\lambda^{-1}{}_k$,使回归系数最小二乘估计的方差明显增大。

总之,代数符号的异常改变,回归系数值的明显变化,以及大的方差是多元共线变量之最小二乘回归系数估计所显示出来的典型性质。

四、岭回归分析

1. 岭迹的计算

从回归分析的理论[1,2]可知,岭回归分析法可以克服最小二乘意义下回归分析所出现的一些缺点。即岭回归的方差比最小二乘估计的方差小,主要是通过估计的偏差来达到,同时,能较准确地估测回归系数。简单地说,就是用$(1+k^\circ)$,其中 $k^\circ>0$,代替自变量相关矩阵 R 中的主对角元素 1,因而,其标准正规方程为:

$$\begin{pmatrix}(1+k^\circ) & r_{12} & \cdots\cdots & r_{1n}\\ r_{21} & (1+k^\circ) & \cdots & r_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ r_{n1} & r_{n2} & \cdots & (1+k^\circ)\end{pmatrix}\begin{pmatrix}\hat{\beta}_1^*\\ \hat{\beta}_2^*\\ \vdots\\ \hat{\beta}_n^*\end{pmatrix}=\begin{pmatrix}r_{1y}\\ r_{2y}\\ \vdots\\ r_{ny}\end{pmatrix} \tag{25}$$

回归系数的岭估计$\hat{\beta}^*=(\hat{\beta}_1^*,\hat{\beta}_2^*\cdots,\hat{\beta}_n^*)$为

$$\hat{\beta}^*(k^\circ)=(R+k^\circ I)^{-1}\tilde{X}'\tilde{Y}=\sum_{k=1}^{n}(\lambda_k+k^\circ)^{-1}c_kP_k \tag{26}$$

且有
$$E(\hat{\beta}^*)=\tilde{\beta}-k^\circ\sum_{k=1}^{n}(\lambda_k+k^\circ)^{-1}P_kP_k'\tilde{\beta} \tag{27}$$

$$Var(\hat{\beta}^*)=\sum_{k=1}^{n}\lambda_k(\lambda_k+k^\circ)^{-2}\sigma^2,P_kP'_k \tag{28}$$

比较(21)和(26)式可知，即使在很小特征值 λ_k 上加一小量 k°，也能大大降低多元共线特征向量的影响。如果 k° 足够小,仅使(26)式中很小的特征值之影响有明显改变,则相应于中等大小和大特征值之特征向量,对 $\hat{\beta}$ 和 $\hat{\beta}^*$ 将有近似相同的影响，因为此时有

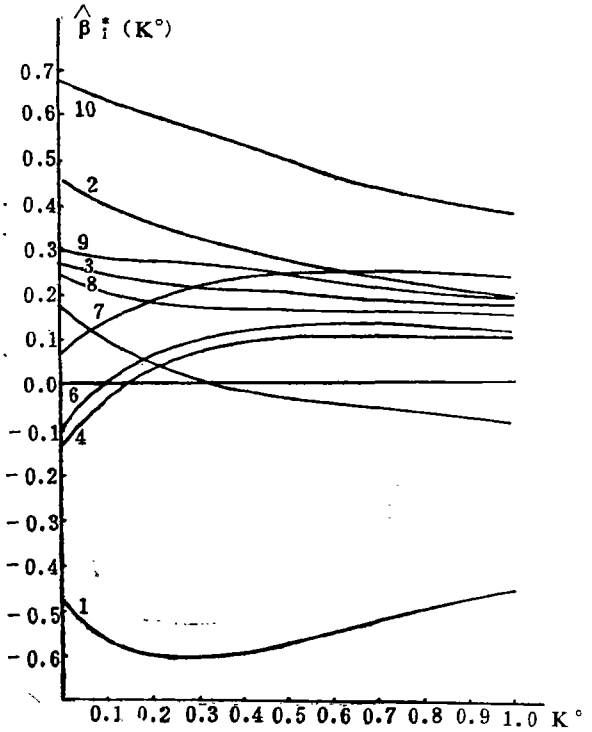

图 1　岭迹图(变量 1—10)

$(\lambda_k+k^\circ)^{-1}\approx\lambda^{-1}{}_k$。由(28)和(24)式看出，岭估计是有偏的，但其方差比最小二乘估计的方差更小，因为对于 $k^\circ>0$，有 $\lambda_k(\lambda_k+k^\circ)^{-2}<\lambda^{-1}_k$。

由表 2 和(26)式求得各个变量在岭参数 k° 由零逐渐增至1的岭回归系数$\{\hat{\beta}_i^*(k^\circ)\}$，然后作出岭迹图$\{\hat{\beta}_i^*(k^\circ)\}\sim k^\circ$，见图1。

2. 岭迹的分析

岭回归的目的是通过适当选择岭参数 k°，而试图大大减小回归系数估计的方差，又只引入小的偏倚。只要 k° 值选择得当，岭回归不仅能减小多元共线性的效应，而且岭估计比最小二乘估计更接近于真实的回归系数。保证岭估计更接近于 $\tilde{\beta}$ 的条件是：

$$0<k^\circ<\sigma_y^2/\tilde{\beta}'\tilde{\beta} \qquad (29)$$

由于上界未知，因此上面不等式仍无法付诸实用。

在 k° 的选择中，有一种称为岭迹(ridge trace)图的方法。利用岭迹来选择 k° 值的依据是：当 k° 增大时，(26)式中系数估计可能有的变化，已在很大程度上排除了小特征值的影响；选择的 k°，要使各回归系数的估计值都能达到稳定。

用岭迹来了解各自变量的作用以及相互关系时，有七种情况[4]：①当$\hat{\beta}_i^*(0)=\hat{\beta}_i^*>0$且较大，从最小二乘法的观点，应将 x_i 看作对 y 有重要影响的因素，若 k° 从 0 开始增加时，$\hat{\beta}_i^*(k^\circ)$显著地下降，而且迅速趋于 0，因而失去“预报能力”，从岭回归的观点看，x_i 对 y 不起重要影响，可去掉这个变量。②在$\hat{\beta}_i^*=\hat{\beta}_i^*(0)>0$，但很小，从最小二乘估计的观点来看，$x_i$ 对 y 的作用不大，随着 k° 略增，$\hat{\beta}_i^*(0)$骤然变为负的，从岭回归观点看，x_i 对 y 有显著影响。③$\hat{\beta}_i^*=\hat{\beta}_i^*(0)>0$，还比较显著，但当 k° 增加时迅速下降且稳定为负值，在最小二乘法中，x_i 是对 y 有“正”影响的显著因素，而在岭回归分析中，x_i 要被看作对 y 有负”影响的因素。④在 $\hat{\beta}_i^*(k^\circ)$和 $\hat{\beta}_j^*(k^\circ)$ 都很不稳定，但其和大体上稳定，这种情况往往发生在自变量间的相关性很大的场合，即 x_i 和 x_j 之间存在多元共线性的场合。因此，从变量选择的观点看，两者只要保存其一就够了。这种情况有助于解释某些回归系数估计的符号不合理。⑤把所有回归系数的岭迹都描在一张图上，如果这些迹线的“不稳定程度”很大，整个系统呈现比较乱的局面，则最小二乘估计不能反映真实的情况。⑥为⑤的情况反过来，则对最小二乘估计可以有更大的信心。⑦有些情况介于⑤和⑥之间，对这种情况，必须适当选择 k° 值。

综上所述，借助岭迹图来选择最佳的变量，应遵循三个规则：(1) 去掉岭回归系数稳

定且绝对值较小的变量；(2) 去掉岭回归系数不稳定但随 $k°$ 的增加迅速趋于零的变量；(3) 去掉一个或若干个具有不稳定岭回归系数的变量。

从图 1 中可得到，变量 x_4, x_5, x_6, x_7 符合规则(3)，应剔除之，x_1 的系数符号变异也应剔除。剩余变量为 $x_2, x_3, x_8, x_9, x_{10}$。

本例的隶属度数值经过了标准化处理后，使得原子集构成了一个正交子集，那么，剩余变量是否构成一个正交子集？为此，我们需绘出 $\sum_{i\in L}\hat{\beta}_i^{*2}(k°)\sim k°$ 图。在几何上，$\sum_{i\in L}\hat{\beta}_i^{*2}(k°)$ 表示岭估计与原点间的平方距离。对一正交子集来说，岭回归系数与原点间的平方距离应为 $\left(\sum_{i\in L}\hat{\beta}_i^{*2}(0)\right)\Big/(1+k°)^2$，其中 $\hat{\beta}_i^{*}(0)=\hat{\beta}_i$，为通常的最小二乘估计。若剩余变量构成一个正交集，则 $\sum_{i\in L}\hat{\beta}_i^{*}(k°)$ 与 $\left(\sum_{i\in L}\hat{\beta}_i^{*2}(0)\right)\Big/(1+k°)^2$ 的图形应大体相同。在此情况下，才可进行下一步分析，否则还须剔除一些变量，直至能产生一个近似正交子集为止。本例的剩余变量 $x_2, x_3, x_8, x_9, x_{10}$ 的系数向量观测平方长度 $\sum_{i\in L}\hat{\beta}_i^{*2}(k°)$ 和正交集得出的期望平方长度 $\left(\sum_{i\in L}\hat{\beta}_i^{*2}(0)\right)\Big/(1+k°)^2$ 的图形见图 2。前者用实线表示，后者用虚线表示。两条曲线十分接近，表明剩余变量构成了一个近似正交子集。

由图 1 可看出，在 $k°=0.2$ 处，岭迹是稳定的，借助隶属度的标准化变量拟合的方程为：

$$\hat{y}=0.3547\,\tilde{x}_2+0.2410\,\tilde{x}_3+0.1738\,\tilde{x}_8+0.2716\,\tilde{x}_9+0.6108\,\tilde{x}_{10} \quad (30)$$

$\hat{y}, \tilde{x}_i$ 均为隶属度的标准化值，由(30)和(13)两式得：

$$\tilde{\mu}'(y)=0.4150\,\hat{y}+0.5151 \quad (31)$$

$\tilde{\mu}'(y)$ 为预测的隶属度。

这样，便从 10 个变量的隶属度数据中，选出了 5 个变量，并建立了岭回归的预测方程。这 5 个变量的回归系数是稳定的，产量与各个因子构呈正相关，与其中的 x_{10}, x_2 的影响最为显著。

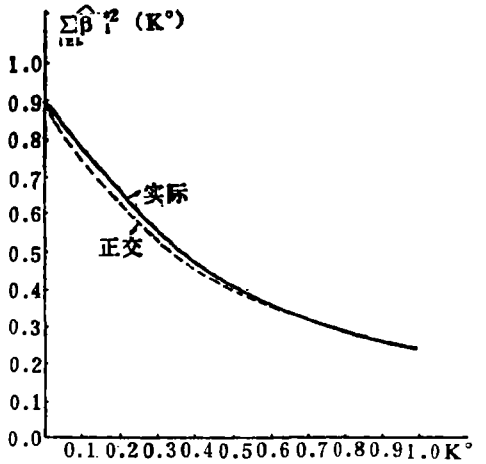

图 2 剩余变量系数向量的平方长度

根据历史拟合情况和(11)式划分的丰、平、欠三级产量年景，规定：

$$\hat{y}'=\begin{cases}\text{丰}, \tilde{\mu}'(y)\geqslant 0.9\\ \text{平}, 0.3<\tilde{\mu}'(y)<0.9\\ \text{欠}, \tilde{\mu}'(y)\leqslant 0.3\end{cases} \quad (32)$$

回代的结果见表 3，历史拟合率为 88.2%。

对 1983 年早稻产量年景预测为平产年，与实况相符。

表 3　1966—1982年早稻产量年景拟合表

年份	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
实况	平	丰	减	减	丰	平	平	减	丰	平	平	平	平	平	丰	减	丰
回代	平	丰	减	减	丰	平	平	减	丰	减	平	平	平	平	丰	平	丰
评定	√	√	√	√	√	√	√	√	√	×	√	√	√	√	√	×	√

五、小　　结

本文将模糊集合嵌入到岭回归分析中，既通过模糊集合来处理早稻产量与各个变量间的模糊性，又在岭回归分析中，对具有多元共线性的隶属度数据进行了处理，选择5个变量来建立回归模型，其岭回归系数接近于真实的回归系数。

(32) 式是根据(11)式中丰、平、欠三级产量年景来确定的，如果需要增加产量年景的预报级别，也可以通过调整(11)式来取得，这就说明了将模糊集合嵌入岭回归方程中，可使模糊集合由定性化向定量化转变。

参 考 文 献

[1] Hoerl, A.E., and R.W. Kennard, Ridge regression applications to nonorthogonal problems, *Technometrics*, 12, 69, 1970.

[2] MCDonald, G.C., and R.C. Schwing, Instabitilies of regression estimates relating air pollution to mortality, *Technometrics*, 15, 463, 1973.

[3] 刘来福，模糊数学及其在生命科学中的应用，生物科学参考资料，第17集，136—148，科学出版社，1983。

[4] 陈希孺等，近代实用回归分析，广西人民出版社，1984。

THE APPLICATION OF THE RIDGE REGRESSION EMBEDDED IN FUZZY SETS TO THE FORECASTING FOR THE ANNUAL RICE YIELD

Li Yaoxian

(Guangxi Zhuang Autonomous Region Observatory)

Abstract

This paper, based on the theories of FUZZY sets and ridge regression, a forecasting equation for the annual early-rice output rank in Yulin district is established.