

# 用带周期分量的逐步回归模型 预测冬小麦产量的试验

王春乙 潘亚茹

(气象科学研究院)

## 提 要

本文首先对粗选因子进行主分量分析,得到主分量组合因子,将其连同预报量和主分量因子的周期分量一起嵌入逐步回归方程,来组建北方主要产麦区冬小麦产量预报模式,收到较好的效果。

通常在应用回归分析方法来预测冬小麦产量时,首先是选取与产量相关显著的环境因子,然后组建一定时效的预报模型。但是,由于受资料样本的限制,使计算的相关系数具有一定的阶段性和不稳定性<sup>[1]</sup>。鉴于此,本文采用带有周期分量的逐步回归方法来组建预测模型<sup>[2]</sup>,使所建的预报模型同时考虑了预报因子、预报因子和预报量的周期分量对回归模型的影响。另外,由于所选取的预报因子性质相近或相同,它们之间往往存在着显著相关,这样会使组建的预测模型稳定性和适应性变差<sup>[3]</sup>。为了解决这个问题,对粗选因子进行主分量分析<sup>[4-5]</sup>。

### 一、产量资料的处理和因子粗选

#### 1. 产量资料处理

据文献[6,7]的研究,农业气象产量预报可用下式表达:

$$y = y_i + y_w + \Delta y \quad (1)$$

式中,  $y$  为农业气象预报产量,  $y_i$  为趋势产量,  $y_w$  为气象产量,  $\Delta y$  为随机噪声,在实际建模过程中,  $\Delta y$  可忽略不计<sup>[8]</sup>。北方主要产麦区(京、津、冀、鲁、豫、陕、晋及苏、皖北部)的冬小麦历年单产用下式计算:

$$y = \frac{\sum_{i=1}^n y_i S_i}{\sum_{i=1}^n S_i} \quad i = 1, 2, \dots, 9 \quad (2)$$

式中,  $y_i$  为某省(市)冬小麦第  $i$  年的实际产量,  $S_i$  为相应的种植面积。通过做冬小麦产量同时间  $t$  (年) 的散布图,可知趋势产量  $y_i$  用二次抛物线来模拟最为理想,表达式为:

$$y_i = a + bt + ct^2 \quad (3)$$

式中,  $t$  为时间(年),  $a$ 、 $b$ 、 $c$  为待定参数。则气象产量为:

$$y_w = y - y_i \quad (4)$$

#### 2. 因子粗选

农作物的生长发育及产量形成直接受周围的生态环境所制约,如光、温、水等主要气象因子,土壤因子,品种、肥料、栽培技术及管理水平等社会生产力因子,它们缺一不可。因此,农作物产量形成是一个复杂的动态变化系统。文献[6,7]认为,土壤因子在短期内变化比较稳定,社会生产力因子可用趋势产量来综合模拟。因而,农作物产量年际间的波动主要受气象因子的影响。赵四强等人<sup>[9-11]</sup>的研究表明,太平洋海温的变化直接影响东亚地区的大气环流,海温作为长期天气过程的重要物理因子,必然与农作物的生长发育乃至产量形成有关。所以,用海温预报农作物产量,是近几年来研究出的一条有效途径,尤其是用它预报大范围农作物产量是可行的。本文利用1949—1981年西太平洋部分海域61个网格点上海温同气象产

量进行单相关普查,粗选出10个相关显著的因子:

$x_1$ , 上年9月20°N、140°E的平均海温;  
 $x_2$ , 上年9月20°N、145°E的平均海温;  
 $x_3$ , 上年10月40°N、145°E的平均海温;  
 $x_4$ , 上年11月45°N、180°E的平均海温;  
 $x_5$ , 上年12月35°N、155°E的平均海温;  
 $x_6$ , 当年1月30°N、175°E的平均海温;  
 $x_7-x_{10}$ , 分别为当年2月35°N、180°, 30°N、150°E, 30°N、165°E, 30°N、180°的平均海温。

粗选因子同 $y_w$ 的单相关系数见表1。

表 1

|       | $x_1$   | $x_2$   | $x_3$   | $x_4$  | $x_5$    |
|-------|---------|---------|---------|--------|----------|
| $y_w$ | -0.4467 | -0.4732 | -0.4085 | 0.5064 | -0.4593  |
|       | $x_6$   | $x_7$   | $x_8$   | $x_9$  | $x_{10}$ |
| $y_w$ | 0.4369  | 0.3910  | 0.4627  | 0.4122 | 0.3991   |

## 二、预报因子、预报量周期分量的生成和粗选因子的主分量分析

### 1. 预报因子、预报量周期分量的生成(2)

设 $F = F_1, F_2, \dots, F_n$ 为时间序列,按试验周期方法,先将资料以周期长度为2分为两组:

$$\begin{array}{cc} F_1 & F_2 \\ F_3 & F_4 \\ \vdots & \vdots \\ F_{n-1} & F_n \end{array}$$

然后求出每列的平均值 $\bar{F}_1, \bar{F}_2$ 。

再按周期长度为3分为3组:

$$\begin{array}{ccc} F_1 & F_2 & F_3 \\ F_4 & F_5 & F_6 \\ \vdots & \vdots & \vdots \\ F_{n-2} & F_{n-1} & F_n \end{array}$$

求出每列的平均值 $\bar{F}_1, \bar{F}_2, \bar{F}_3$ ,

如此反复进行下去,直到分组个数不大

于资料样本的一半 $\left(\frac{n}{2}\right)$ 为止。则最后1组为:

$$\begin{array}{ccc} F_1 & F_2 & \dots F_{\frac{n}{2}} \\ F_{\frac{n}{2}+1} & F_{\frac{n}{2}+2} & \dots F_n \end{array}$$

求出每列的平均值 $\bar{F}_1^{(\frac{n}{2})}, \bar{F}_2^{(\frac{n}{2})}, \dots, \bar{F}_{\frac{n}{2}}^{(\frac{n}{2})}$ 。以各周期的平均值构成的周期序列,本文定义为周期分量。

### 2. 粗选因子的主分量(12)

利用粗选的10个因子构成原始资料矩阵 $x_{ij}$  ( $i=1, 2, \dots, m$ ,  $m$ 为因子数;  $j=1, 2, \dots, n$ ,  $n$ 为时间),对 $x$ 进行中心化处理:

$$z_{ij} = x_{ij} - \bar{x}_i \quad (5)$$

$$\bar{x}_i = \sum_{j=1}^n x_{ij} / N$$

进而求出矩阵 $x$ 的距平矩阵 $Z$ ,相应的协方差矩阵 $S$ 为:

$$S = Z \cdot Z^T \quad (6)$$

其中, $Z^T$ 为 $Z$ 的转置矩阵。则主分量矩阵 $x'$ 为

$$x' = U \cdot x \quad (7)$$

式中,矩阵 $U$ 为矩阵 $S$ 的特征向量矩阵。在建模时,把主分量矩阵中的每一行向量作为新的预报因子(主分量因子),记为 $x'_i$ 。

## 三、气象产量模式的建立和产量预报模式的验证

### 1. 气象产量模式的建立

为了便于比较和分析,本文共建4个模式,详见框图(图1)。模式1—4的历史拟合和回代预报结果如图2所示。

通过以上分析,模式4性能较好,它的历史拟合及预报均比较稳定。

### 2. 产量预报模式的验证

我们使用1951—1985年海温和气象产量

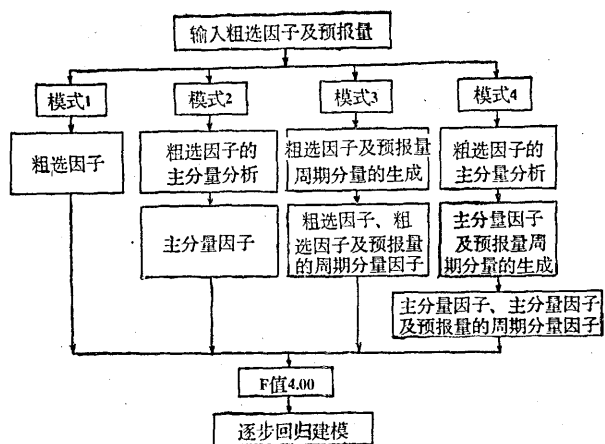


图1

资料, 利用模式4已确定的5个变量组建的气象产量多元回归模型如下:

$$y_w = 9.12 + 0.55F_9 + 0.61F_5 - 0.69x_1' - 0.38F_{14} + 0.31F_{10} \quad (8)$$

方程的复相关系数为0.9302, 剩余方差为2.47。趋势产量可用(3)式外推得出。则1986、1987两年北方9省(市)冬小麦产量的外推预报结果如表2所示。从表2可知, 1986、1987两年预测的相对误差均在6%以下, 预报趋势基本准确, 表明该模式可以投入业务使用, 并且具有较高的可信程度。

表2 1986、1987年冬小麦产量预测结果

| 类别<br>年 | 实际产量  | 预测产量  | 绝对误差 | 相对误差  |
|---------|-------|-------|------|-------|
| 1986年   | 460.0 | 474.4 | 14.4 | 3.12% |
| 1987年   | 453.1 | 471.3 | 18.2 | 4.01% |

模式1—4的方程及其复相关系数、剩余方差见表3。

#### 四、结 论

通过本文的分析表明, 在粗选的众多预报因子彼此之间存在多元共线性的情况下, 首先对粗选因子进行主分量分析, 再把预报量和主分量因子的周期分量一同嵌入逐步回

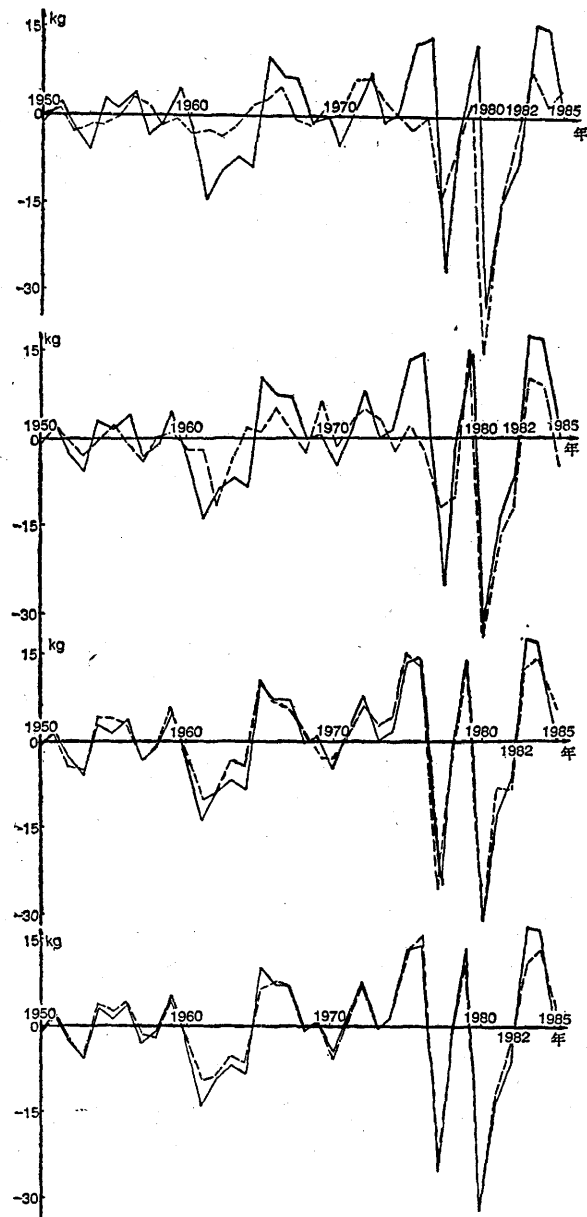


图2

图中: 实线为实际产量值, 虚线为拟合产量值; 1982—1985年为模式预测值; a、b、c、d 分别为模式1、2、3、4的结果

归, 来挑选因子组建预报模型, 是一种较好的处理方法。该方法物理意义较为明确, 具有一定的普遍意义。

表 3

各方程及复相关系数(R)、剩余方差(S)

|      | 方 程                                                                         | R      | S    |
|------|-----------------------------------------------------------------------------|--------|------|
| 模式 1 | $Y_w = 13.88 - 0.37x_3 + 0.56x_4$                                           | 2.6591 | 7.96 |
| 模式 2 | $Y_w = -1.11 - 0.47x_1'$                                                    | 0.6758 | 7.67 |
| 模式 3 | $Y_w = 9.43 - 0.39x_1 + 0.22F_9 + 0.69F_5 - 0.28F_8 - 0.14F_{12} + 0.42x_6$ | 0.9082 | 3.83 |
| 模式 4 | $Y_w = 8.61 + 0.59F_9 + 0.56F_5 - 0.79x_1' - 0.21F_{14} + 0.16F_{10}$       | 0.9282 | 2.65 |

注:  $x_1$  为第一分量因子;  $F_9$ 、 $F_5$  分别为预报量的 9、5 年周期分量;  $F_8$ 、 $F_{12}$  分别为第四组选因子的 8、12 年周期分量;  $F_{14}$ 、 $F_{10}$  分别为第一分量因子的 14、10 年周期分量。

### 参考文献

- (1) 朱盛明, 相关系数稳定性分析方法及其应用, 气象学报, 第 40 卷, 第 4 期, 497—501, 1982.
- (2) 魏凤英等, 带有周期分量的多元逐步回归, 气象科学研究院院刊, 第 1 卷, 第 1 期, 94—101, 1986.
- (3) 森信成, 使用数值预报产品制作 192 小时降水预报 (经验正交函数在多元回归方程中的应用), 山西气象科技 (译文) 第 3 期, 64—71, 1985.
- (4) Isebrands, J.G., Thomas R.C., Introduction to use and interpretation of principal component analysis in forest biology, USDA Forest service, general technical report NC-17, 1—19, 1975.
- (5) Louis, T.S., Sharon, K.L. and D.M. James., Principal component of larle scage general circulation features interpretation and use in climate/crop yield models, Fifth Conference on probability and Statistic in Atmospheric sciences, 16—21, (9).
- (6) 王馥棠等, 产量预报方法研究 (一)——总论与趋势产量, 气象科学技术集刊, 第 3 期, 60—69, 1982.
- (7) 王馥棠, 产量预报方法研究 (二)——气象产量, 农业气象预报文集, 气象出版社, 1982.
- (8) Claupe, Duchon, Corn yield prediction using climatology Journal of Climate and Applied Meteorology Vol. 25, No. 5, 375—402, 1986.
- (9) 赵四强, 应用海温预报粮食产量的初步探讨, 科学通报, 第 27 卷, 第 30 期, 1269—1272, 1982.
- (10) 赵四强, 欧亚大气环流和北太平洋海温与我国粮食产量, 农业气象, 第四卷, 第三期, 1—5, 1983.
- (11) Baier W., Crop-weather models and their use in yield assessments, WMO Tec. Note, No. 151, 1977.
- (12) 阳含熙等, 植物生态学的数量分类方法, 232—245, 科学出版社, 1983.

## An approach to winter wheat yield prediction using a model with orthogonal factors and periodic components

Wang Chunyi Pan Yaru

(Academy of Meteorological Science)

### Abstract

Since the prediction model is influenced by the multicollinearity of the selected factors and the unstability of the correlated coefficient, principal component analysis is used in this paper. Along with the compound factors of principal components of primary selected factors and the periodic components of prediction factors as well as the principal components itself, a stepwite regression model for predicting winter wheat yield in the major winter wheat areas of North Chian is established. Both its fitness and feasibility are better than those of the ordinary stepwise regression models.