

# 逐步回归自相关分析

杨 昕<sup>1)</sup>

(中国气象科学研究院,北京 100081)

## 提 要

作者从时间序列的自相关分析出发,提出了一种适于回归统计分析的生成预报因子的简单方法,建立起自相关预报模型。实例计算表明,此方法不仅预报精度较高,而且在对周期性相关因子的选择上具有等同的统计、检验前提。避免了一般方法因周期长度不同而带来统计、检验差异的影响。因而得到的回归因子较客观、真实。另外,还具有一定的分析时间序列隐含周期的能力。

**关键词:** 时间序列 逐步回归 自相关分析

### 1 问题的提出

在气象问题中,所处理的大多数时间序列,一般都是非平稳的,因此,当用平稳随机序列作预报时,由于周期性的影响,自相关系数就不符合各态历经的基本条件。因此,需要对此进行有关处理才能有较好的预报效果,具体问题不同,处理方式也各不一样<sup>[1,2]</sup>。如何对含有周期性波动的非平稳时间序列进行客观有效的分析,对提高预报精度具有重要意义。

本文提出了一种较为实用的时间序列分析方法。它主要是对序列中各元素与相隔等时间距离的前期元素进行整体相关性统计分析来建立回归预报模型,此方法预报因子生成方式简单,且各因子与预报量可在相同的统计、检验前提下进行因子的筛选,所获得的因子较客观。而且此方法不仅有预报功能,还有一定的分辨序列中隐含周期的能力。做为实例应用,本文还给出与其它几种预报方法的效果比较和对汉口站汛期降水量的周期分析情况,结果令人满意,因此可做为一种实用的分析预报工具。

### 2 预报因子的生成及预报方程

设一气象要素时间序列  $X(t)$ :

$$X(t) = \{x(1), x(2), \dots, x(n)\} \quad (1)$$

其中  $n$  为样本数。

设后延时间为  $i$  个单位的预报因子为自相关序列  $X_i$ :

$$X_i = \{x(1), x(2), \dots, x(n-i)\} \\ i = 1, 2, \dots, m \quad (2)$$

其中,  $m = \text{INT}(n/2)$ , 为  $n/2$  的整数部分。

为进行回归统计计算,取以上  $m$  个预报因子及预报量  $X(t)$  的公共对应部分,得到如下参加计算的样本:

$$X_j = \{x(m-j+1), x(m-j+2), \\ \dots, x(n-j)\} \quad j = 1, 2, \dots, m \quad (3) \\ X = \{x(m+1), x(m+2), \dots, x(n)\} \quad (4)$$

由此,可建立线性自回归预报方程:

$$\hat{X} = \alpha_0 + \sum_{j=1}^m \alpha_j \beta_j + e \quad (5)$$

其中,  $\hat{X}$  为预报计算值,  $\alpha_0$  为回归常数,  $\alpha_j$  为因子回归系数,  $\beta_j$  为预报因子  $X_j$  所对应的实测值,  $e$  为白噪声。

利用逐步回归技术来估计系数  $\alpha_0, \alpha_j$ 。给定  $F$  值下,如  $\alpha_j$  不等于零,则表示原时间序列中存在相隔距离为  $j$  个时间单位的自相关性。显然,这种相关性可视为一种周期规律。只不过这种周期性规律与通常意义上由方差分析等方法所得到的周期性结论在含义上有所不同。它是以等时间间隔上两两相邻的样本来进行方差贡献的统计检验的,故这种相关性是可正可负的。参加预报的也都是临近的周期性相关实测值,这一点也符合客观实际情况。

### 3 实例应用及效果比较分析

1) 94 级研究生。

我们对江苏徐圩盐场气象站 1952—1981 年共 30 年 ( $n=30$ ) 的 6—8 月降水资料进行了分析计算, 计算中取  $m=15, F=2.0$ , 得到预报方程为:

$$\hat{X} = 430.4 - 0.31\beta_2 + 0.37\beta_{14}$$

$\beta_2, \beta_{14}$  分别对应 1980 年及 1968 年的实测值,

则得到 1982 年的预报值为 489mm, 实测为 508mm, 仅相差 19mm。

表 1 给出了利用本文所述方法计算的 1982—1989 年预报及实测值。同时还给出了传统自相关分析方法、逐步回归周期分析方法及周期外推分析方法的计算结果。

表 1 4 种方法的预报效果比较(单位:mm)

年份	1982	1983	1984	1985	1986	1987	1988	1989	$ \bar{E} $
实测值	508	515	523	420	585	453	365	528	
本文方法	489	450	498	527	468	466	481	536	0.13
传统自相关分析	633	669	588	463	668	590	566	598	0.25
逐步回归周期分析	457	547	611	626	450	528	600	635	0.26
周期外推分析	567	630	683	755	498	422	603	738	0.33

注:  $E = (\text{预报值}/\text{实测值}) - 1$ ;  $|\bar{E}|$  为 8 年绝对值平均

从表中看出本方法的预报效果最好。如以 550mm 为气候平均值, 则 8 年中有 7 年的趋势预报是正确的。特别象 1987、1988 年连续两年降水量极少年都能有很好地反映, 这是其它方法所不及的。

为验证本方法在因子选择上具有长短周期因子等机会入选的特性, 我们利用汉口站 1880—1969 年共 90 年 ( $n=90$ ) 的汛期(6—9 月)降水量资料进行了分析计算, 计算中取  $m=45$  年, 并与逐步回归周期分析方法进行了对比, 结果见表 2。

表 2 周期性因子的选取及对比

方法	检验信度	入选因子的周期性长度/年		
本文方法	$F=3.8$	8	22	39
对比方法	$F=2.6$	42	43	44

从表 2 可见, 在控制信度要求下所入选的 3 个周期性因子中, 本文所述方法的周期长度分布较为合理, 短、中、长期均有人选, 而且, 特别是象具有明显物理意义的 22 年周期, 也能很好地反映出来, 且是第一显著因子。而对比方法由于其在回归预报因子的生

成上所产生的回归信度要求的不等性<sup>[3,4]</sup>从而导致只有长周期入选。这表明, 在把时间序列转化为多元回归分析时, 因子的生成方式十分重要, 即各因子的回归前提必须尽可能一致才不会产生客观信息的改变。

#### 4 小结

本文所述方法是将时间序列问题转化为多元回归问题进行处理的另一尝试。同时也为将时间序列分析更好地运用到长期天气预报, 提供了一种可行性途径。

本方法尚有待进一步探讨, 如它的理论分析及对隐含周期分辨方面还需做一些深入工作。

#### 参考文献

- 1 史久恩, 徐群. 长江中下游夏季降水长期预报的初步探讨. 气象学报, 1974, 32(2).
- 2 杨昕. 降水量的年际波动特点及单变量自相关统计预报. 安徽气象, 1992, (3).
- 3 魏凤英, 赵溱, 张先恭. 逐步回归周期分析. 气象, 1983, 9(2).
- 4 魏凤英, 张先恭, 曹鸿兴. 逐步回归周期分析的改进方案及其在气候预测中的应用. 气象, 1989, 15(7).

## Lag Correlation Analysis by Stepwise Regression Method

Yang Xin

(Chinese Academy of Meteorological Sciences, Beijing 100081)

### Abstract

A new lag correlation analysis model is developed with a stepwise regression method. This model gives a simple method to generate predictors which suit to multivariate regression. Computational results show that this model is more objective in periodic predictor selection than general models. All predictors have the same premise of statistic and significance test to predictant. What's more, this model has some resolving power of hidden periodicity.

**Key Words:** time series stepwise regression lag correlation analysis