

逐步回归周期分析与多层递阶 方法相结合的时间序列分析预报模型*

李 邦 宪

(浙江省金华市气象台)

逐步回归周期分析^[1]是针对方差分析的不足而提出的一种时间序列分析方法, 作者^[2]将其应用于长期预报, 取得了一定的效果, 但也存在着某些缺陷。其主要表现在实际应用时, 随着预报时效的增长, 预报效果明显趋于不稳定。这里主要有两个方面的原因: 一是某些时间序列本身的周期性不强, 所选的周期显著性水平低, 此时则不宜单纯用周期分析外推做长期预报, 对此, 作者已在另文中作过研究^[3,4]。二是在逐步回归周期分析预报模型中, 回归系数取为固定的参数, 而天气系统则是时变系统。这样, 若用固定参数预报模型来描述一个时变参数系统的状态, 其拟合率虽然可以很高, 但随着预报时效的增长, 则预报效果趋于不稳定是必然的。本文针对这个问题, 提出了一种逐步回归周期分析与多层递阶方法相结合的时间序列分析预报模型, 实际使用效果较好。

1. 思路及预报模型

逐步回归周期分析与多层递阶方法相结合的基本思路是: 先用逐步回归周期分析提取显著周期分量, 并将诸个周期分量视为预报因子, 然后利用多层递阶方法^[5], 将天气系统的状态预报分离成两部分, 即对系统的时变参数进行预报和在此基础上对系统的状态进行预报。

将统计预报数学模型取为:

$$y(t) = \sum_{i=1}^k \beta_i(t) f_i(t) + e(t) \quad (1)$$

其中 $y(t)$ 为预报量, $f_i(t)$ 为 $y(t)$ 的显著周期函数, $\beta_i(t)$ 为时变参数, $e(t)$ 为白噪声。 $f_i(t)$ 利用逐步回归周期分析求取, $\beta_i(t)$ 则可通过多层递阶方法计算。

2. 计算步骤

1) 首先仿照方差分析求试验周期的方法, 将预报对象序列 $y(t)$ 依次按长度 $l (2 \leq l \leq m)$ 进行分组, 然后对各组求平均, 则可得到 $m-1$ 个试验周期序列。其中 m 为不大于 $\frac{n}{2}$ 的最大整数, n 为序列样本长度。并将这 $m-1$ 个试验周期序列视为候选因子 $x_1, x_2, \dots, x_{m-1}$ 。

2) 给定显著性检验水平 F_0 , 按照逐步回归计算步骤^[6], 对 $x_1, x_2, \dots, x_{m-1}$ 进行因子的逐个引入或剔除, 直到既无因子可剔除又无因子可引入为止。显然, 若共筛选出 k 个因子, 记作 $f_1(t), f_2(t), \dots, f_k(t)$, 则表明该时间序列存在 k 个显著周期。

3) 令 $\varphi^T(t) = [f_1(t), f_2(t), \dots, f_k(t)]$
 $\theta^T(t) = [\beta_1(t), \beta_2(t), \dots, \beta_k(t)]$

则 (1) 式可写成: $y(t) = \varphi^T(t) \theta(t) + e(t) \quad (2)$

引进时变参数递推算公式:

*本文于 1987 年 6 月 3 日收到, 1988 年 4 月 4 日收到最后修改稿。

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{\|\varphi(t)\|^2} \varphi(t) \cdot [y(t) - \varphi^T(t) \hat{\theta}(t-1)] \quad (3)$$

由历史资料,则可得到参数估值序列 $\{\hat{\theta}(t)\}$ 。分析参数估值序列 $\{\hat{\theta}(t)\}$ 的演变情况,根据其不同变化特点,通过适当的数学手段,建立其所满足的参数预报模型。常用的有均值近似法、多层 AR 递阶模型、分段周期变量法等。

4) 最后在参数预报的基础上,运用 $y(t)$ 的向前一步预报公式:

$$\hat{y}(n+1/n) = \varphi^T(n+1) \hat{\theta}^*(n+1) \quad (4)$$

即可得到所要求的预报量。其中 $\hat{y}(n+1/n)$ 为 $y(n)$ 的向前一步预报值, $\hat{\theta}^*(n+1)$ 为 $\hat{\theta}(n)$ 的向前一步预报值。

3. 实例计算分析

以 1953—1982 年金华市 6 月份降水量作为预报对象 $y(t)$, 取信度检验值 $F_0 = 10.0$, 利用逐步回归周期分析方法分别提取 10 年、11 年和 15 年三个显著周期分量, 其 F 值分别达到 43.8、30.3 和 10.7, 为极显著水平。

时变参数初值取为零, 应用时变参数递推算法公式(3), 确定参数估值序列 $\{\theta(t)\}$, 见表 1。

表 1 时变参数 $\beta_1, \beta_2, \beta_3$ 的估值

序 号	年 份	$\beta_1(t)$	$\beta_2(t)$	$\beta_3(t)$	序 号	年 份	$\beta_1(t)$	$\beta_2(t)$	$\beta_3(t)$
1	1953	0.33176	0.27011	0.26458	16	1968	0.34033	0.31408	0.40195
2	1954	0.50210	0.48938	0.50319	17	1969	0.33464	0.30941	0.39071
3	1955	0.44950	0.44090	0.44507	18	1970	0.34405	0.31890	0.40751
4	1956	0.26526	0.26049	0.23042	19	1971	0.40639	0.39476	0.47251
5	1957	0.21609	0.23217	0.19963	20	1972	0.27053	0.28073	0.39163
6	1958	0.22585	0.24329	0.21495	21	1973	0.36342	0.39833	0.47291
7	1959	0.33546	0.35673	0.31442	22	1974	0.17027	0.25278	0.34859
8	1960	0.32335	0.34127	0.29961	23	1975	0.28305	0.32428	0.43393
9	1961	0.31370	0.33161	0.28827	24	1976	0.29294	0.34292	0.44692
10	1962	0.34891	0.37641	0.33356	25	1977	0.30588	0.35911	0.46443
11	1963	0.39002	0.41023	0.36499	26	1978	0.20238	0.22348	0.32247
12	1964	0.22159	0.28463	0.27749	27	1979	0.22906	0.24225	0.34201
13	1965	0.27905	0.34751	0.31006	28	1980	0.12669	0.16047	0.24827
14	1966	0.18870	0.20418	0.24700	29	1981	0.21320	0.24365	0.30232
15	1967	0.37560	0.34854	0.45241	30	1982	0.28553	0.31739	0.37788

从表 1 可以看出, $\{\theta(t)\}$ 年际变化幅度较大, 是一个时变参数序列。考虑到 $f_i(t)$ 是周期变量, 时变参数序列具有明显的自相关特性, 因此, 参数预报采用选点法 AR 模型, 即取 10, 11 和 15 年三阶 AR 预报模型:

$$\hat{\theta}_i^*(t) = \varphi^T(i, t) a_i + e(t) \quad (5)$$

其中

$$\varphi^T(i, t) = [\hat{\theta}_i(t-10), \hat{\theta}_i(t-11), \hat{\theta}_i(t-15)]$$

$$a_i = [a_i(10), a_i(11), a_i(15)]$$

a_i 为时变参数的自回归系数, 利用(5)式则可求出 1983—1986 年的参数预报值(见表 2), 然后在参数预报的基础上, 应用(4)式即可求出 $y(t)$ 的预报值(见表 3(A)栏)。

考虑到太远的数据传递的信息已失去了应有的价值, 反而会起到某些干扰作用, 本文采用有限记忆法, 固定样本数 $n=10, 11, \dots, 20$, $y(t)$ 每向后推一年, 就舍去一个最前面的样本。比较不同固定样本数的计算结果, 发现 $n=15$ 时预报效果最为稳定(见表 3(B)栏)。

表 2 时变参数预报值

年份	$\beta_1^*(t)$	$\beta_2^*(t)$	$\beta_3^*(t)$
1983	0.29784	0.31400	0.35232
1984	0.29878	0.31522	0.35421
1985	0.29851	0.31504	0.35138
1986	0.29390	0.31207	0.34971

表 3 3 种不同方法的预报效果比较

年份	实测值	预报值(y)			残差(y-y)			相对误差(%)		
		A	B	C	A	B	C	A	B	C
1983	245	257	267	215	12	22	-30	4.9	9.0	12.2
1984	325	328	306	433	3	19	108	0.9	5.8	33.2
1985	168	294	307	374	126	139	206	75.0	82.7	122.6
1986	177	232	161	194	55	-16	17	31.1	9.0	9.6

为了便于比较,取相同的资料建立了逐步回归周期分析预报模型:

$$y = -189.4 + 0.4952x_{10} + 0.5344x_{11} + 0.7801x_{15}$$

用该方程计算的 1983—1986 年预报值列在表 3 (C) 栏。从表 3 中可以看出,多层递阶方法能显著减少逐步回归周期分析的预报误差。

4. 讨论和结论

1) 逐步回归周期分析对时间序列的隐含周期有较高的分辨率;而多层递阶方法则充分考虑了系统的时变特性,因而,两者相结合可以有效地提高时间序列分析的预报效果。

2) 本文实际计算表明,采用有限记忆法可以提高预报效果。但应该指出,本文发现 $n=15$ 时预报效果最佳,是在计算了 $n=10, 11, \dots, 20$ 的有限记忆法与实测值比较后作出的,并没有在理论上加以说明,它是否具有普遍意义,还有待于更深入的探讨。

3) 计算时一般入选周期应控制在 4 个以下, F 信度检验水平应适当提高,这样有利于预报效果的稳定。如果为了追求拟合率而降低信度要求,往往会导致预报失败。

4) 本方法计算工作量较大,手工操作计算一般难以完成,需借助电子计算机才能投入实际业务工作中使用。

参 考 文 献

- [1] 魏凤英、赵溱、张先恭, 逐步回归周期分析, 气象, 1983, 第 3 期, 2-4。
- [2] 李邦宪, 试用逐步回归周期分析作月降水预报, 浙江气象科技, 7, 4, 11-12, 1985。
- [3] 李邦宪, 逐步回归双重分析, 气象, 13, 10, 40-42, 1987。
- [4] 李邦宪, 逐步回归双重分析及其在长期天气预报中的应用, 大气科学, 13, 3, 373-377, 1989。
- [5] 韩志刚等, 黑龙江省冬季平均气温的多层递阶长期预报模型, 大气科学, 9, 2, 171-177, 1985。
- [6] 冯士雍, 回归分析方法, 117-128, 科学出版社, 1985。

A TIME SERIES FORECASTING METHOD COMBINING THE STEPWISE REGRE- SSION PERIOD ANALYSIS WITH THE RECURSIVE ESTIMATE METHOD

Li Bangxian

(Meteorological Observatory of Jinhua City, Zhejiang Province)

Abstract

The paper is aimed at the defect of the stepwise regression period analysis, a improved time series forecasting model is presented. It combines the stepwise regression period analysis with the recursive estimate method, and is apply ed to the long-range forecasting of the precipitation on June at Jinhua City. The result shows that the improved method is satisfying.