第13卷 第3期
1989年 9月

大 气 科 学
SCIENTIA ATMOSPHERICA SINICA

Vol. 13, No. 3
Sept., 1989

逐步回归双重分析及其在长期天气预报中的应用

李邦宪

(浙江省金华市气象台)

提 要

本文针对多元分析和周期分析各自存在的缺陷，提出了一种新的统计预报方法——逐步回归双重分析方法，它利用逐步回归技术将因子筛选和周期分析有机结合，并反映在同一预报方程中.

本文以金华市5月份降水量长期预报为例，进行了初步的分析和尝试，表明该方法与多元分析和周期分析比较，其拟合误差与预报误差均较小，预报效果稳定，对长期天气预报有较好的实用价值.

一、问题的提出

多元分析和时间序列分析是统计预报的两大分支，近年来发展较快，为目前基层气象台站制作长期天气预报的主要方法之一.

多元分析假设气象要素的变化主要受前期因子的支配，而时间序列分析则强调气象要素的变化取决于它自身的演变规律. 显然，这两种方法各有侧重，但都不全面. 李道松[1]、泮仰耕等[2]针对多元分析和时间序列分析各自存在的缺陷，提出了一种回归分析与残差周期迭加相结合的预报方法，他们先用回归分析方法建立预报方程，然后将实测值减去回归方程的拟合值，产生一残差序列，用方差分析方法对该残差序列提取显著周期分量，最后将这两部分值迭加作为预报值，取得了一定的效果. 但问题在于，用回归分析方法建立预报方程时，并没有把周期波从原序列中分离出去，使得回归计算时易产生虚假拟合，因而影响了回归方程的预报效果；而回归方程的拟合残差序列与预报的残差也不一定属于同一分布，使得残差的方差分析订正效果也不太稳定. 事实上前期因子和周期性往往同时支配着气象要素的变化，为此，作者提出了一种利用逐步回归技术同时进行因子筛选和周期分析的方法——逐步回归双重分析，经实际业务工作使用检验，表明该方法具有良好的预报效果，对基层气象台站开展长期天气预报有较高的使用价值.

二、思路及统计数学模型

如前所述，气象要素的变化既受前期因子的支配，又同时存在自身的演变规律. 因此，本文将统计模型取为:

1987年2月1日收到，6月8日收到修改稿

$$y(t)=\sum_{i=1}^{K_1} a_i f_{1i}(t)+\sum_{i=1}^{K_2} b_i f_{2i}(t)+\varepsilon(t)$$

其中 $y(t)$ 为预报对象，$f_{1i}(t)$ 为预报因子，$f_{i2}(t)$ 为预报对象的周期分量，$\varepsilon(t)$ 为随机误差，a_i、b_i 为权重系数，k_1、k_2 分别为预报因子和周期分量个数. 在建立预报方程时不考虑随机误差，也即视 $\varepsilon(t)$ 为预报误差. 在实际工作中应使 $\varepsilon(t)$ 尽可能小.

在上述数学模型中，关键是要解决两个问题：一是预报对象的周期分量 $f_{2i}(t)$ 如何提取，二是权重系数 a_i、b_i 如何确定. 魏凤英等[3]提出的逐步回归周期分析弥补了方差分析的不足，作者[4]将此方法应用于实际业务预报，表明用该方法提取的周期分量明显优于方差分析，效果较好，且能在提取显著周期分量的同时确定各周期分量的权重系数. 具体做法是将周期分量 $f_{2i}(t)$ 取为预报对象序列 $y(t)$ 的分组平均值序列，然后用逐步回归方法提取显著周期并确定权重系数 b_i. 而前期因子的筛选和权重系数 a_i 的计算也能用逐步回归方法加以解决. 因此，我们自然可以利用逐步回归技术同时进行预报因子的筛选和时间序列周期分析，并确定出各自的权重系数 a_i 和 b_i，从而建立起预报方程. 我们将该方法称之为逐步回归双重分析.

三、计 算 方 法

1. 计算试验周期序列

首先按方差分析求试验周期的方法[5]，将时间序列 $y(t)$ 依次按长度 $l(2\leqslant l\leqslant m)$ 进行分组：

$$\begin{aligned}
&y(1),\ \cdots\cdots\cdots,\ y(i),\ \cdots\cdots\cdots\cdots,\ y(l)\\
&y(1+l),\ \cdots\cdots\cdots,\ y(i+l),\ \cdots\cdots\cdots\cdots,\ y(2l)\\
&\cdots\cdots\\
&y[1+(n_0-1)l],\cdots\cdots,\ y[i+(n_0-1)l],\cdots\cdots,\ y(n)
\end{aligned}$$

其中 n 为时间序列 $y(t)$ 的样本长度，n_0 为满足 $i+(n_0-1)l\leqslant n$ 的最大整数，m 为不大于 $n/2$ 的最大整数.

对以上各组分别求平均，即

$$f_k(i)=\frac{1}{n_0}\sum_{j=1}^{n_0} y[i+(j-1)l]\qquad (i=1,2,\cdots\cdots,\ l;\ k=2,3,\cdots\cdots,\ m)$$

则可得到一个样本长度为 l 的平均值序列，称之为周期长度为 l 的试验周期序列. 按不同长度分组可得到 $(m-1)$ 个试验周期序列. 将各试验周期序列按照周期性外延，使其各试验周期序列的样本长度均为 n，并将这 $(m-1)$ 个新序列视为预报因子 $x_2, x_3, \cdots, x_m$，将时间序列 $y(t)$ 记为 x_1.

2. 粗选前期因子

在给定信度水平下，对大量的有一定物理意义的前期因子进行相关普查，选取相关系数绝对值 $|R|\geqslant R_a$ 的 k 个前期因子，记为 $x_{m+1}, x_{m+2}, \cdots\cdots, x_{m+k}$.

由于在实际业务工作中，候选因子常多达上千个，因而粗选因子是必要的. 实践表明，经过这一步粗选，不仅可以大大减少逐步回归的计算量，而且有利于提高预报方程的稳定性，关于这个问题我们将在另文中分析.

3. 逐步回归计算

按照逐步回归计算步骤[6]，在给定的F检验水平下，对 $x_2, x_3, \cdots\cdots, x_{m+k}$ 进行变量的引入或剔除，直到既无变量可剔除又无变量可引入为止，记下被选变量的序号i. 计算各被选变量的回归系数，建立预报方程. 显然，序号 $i \leqslant m$ 时，则该变量就是预报对象序列所含的显著周期分量；序号 $i > m$ 时，则该变量为前期因子. 因篇幅所限，逐步回归的计算步骤在此不作详述，读者可参考有关数理统计文献 [7].

四、实例计算分析

1. 资料选取

前期因子取1956 — 1983年各月北太平洋37个格点的月平均海温、500hPa环流特征量和高度场网格点资料及单站气象要素资料，预报对象为金华市5月份降水量.

2. 相关普查

为能较早地发布预报，1月份前期因子资料取当年相关，其余各月取隔年相关. 经计算机普查，选取了相关系数绝对值 $|R| \geqslant 0.4$ 的22个前期因子.

3. 建立预报方程

由计算机根据预报量资料自动计算出各个试验周期序列，与22个前期因子一起进行逐步回归计算，取F＝5时，选入6个变量，方程复相关系数达0.97. 预报方程如下:

$$y = -345.0 + 0.777x_9 + 0.635x_{10} + 0.597x_{11} - 6.90x_{25} + 216.4x_{27} + 536.4x_{31}$$

其中前三个变量分别为9、10和11年周期分量，x_{25} 为上年2月(20 °N, 120 °E)与(35 °N, 135 °E)两个格点的500hPa高度差，x_{27} 为上年3月亚欧地区经向环流指数，x_{31} 为上年7月亚欧地区经向环流指数.

该方程经1984, 1985两年独立样本试报检验和1986年实际业务使用，表明预报效果较为稳定. 方程的拟合和预报效果分别见下节表1和表2.

五、与逐步回归、周期分析的比较

为了与通常的逐步回归和周期分析的预报效果进行比较，我们用相同的因子资料和F检验值进行逐步回归计算，取F＝5，选入3个因子，复相关系数为0.86，回归方程为:

$$y = 232.3 - 415.6x_1 + 516.5x_2 + 3.08x_3$$

其中 x_1 为上年7月亚欧地区纬向环流指数，x_2 为上年7月亚欧地区经向环流指数，x_3 为金华市上年11月中旬降水量.

考虑到周期分析资料样本长度太短易引起预报效果不稳定，为此，用1953 — 1983共31年资料进行逐步回归周期分析计算，取F＝5，选入5年、7年、10年和15年4个周期分量，复相关系数为0.94，预报方程为:

$$y = -203.0 - 0.983x_5 + 0.803x_7 + 1.126x_{10} + 0.950x_{15}$$

以上两种方法的拟合和试报效果也分别列在表1和表2.

从表1和表2中可以看到，逐步回归双重分析的拟合和预报效果都明显优于其它两

表1 三种不同方法的拟合情况

年份	实测值	双重分析	逐步回归	周期分析	年份	实测值	双重分析	逐步回归	周期分析
1957	163	132	163	206	1971	273	328	284	268
1958	440	434	450	458	1972	153	144	137	72
1959	200	191	109	180	1973	516	472	421	500
1960	175	173	187	168	1974	218	219	216	244
1961	286	289	278	247	1975	215	192	231	209
1962	262	256	212	220	1976	120	90	135	148
1963	229	209	147	221	1977	370	388	338	426
1964	218	234	284	219	1978	96	138	148	123
1965	159	143	207	166	1979	174	176	205	154
1966	123	122	214	114	1980	155	211	156	168
1967	362	347	283	279	1981	137	126	175	148
1968	188	196	202	172	1982	82	116	85	160
1969	223	208	148	255	1983	197	237	256	171
1970	232	194	296	254	平均拟合绝对误差		20.4	38.7	27.4

表2 三种不同方法的试报及预报情况

年 份	实测值	双重分析		逐步回归		周 期 分 析	
		预报值	误差值	预报值	误差值	预报值	误差值
1984	169	164	−5	112	−57	412	243
1985	125	220	95	311	186	196	71
1986	140	154	14	81	−59	234	94
平均预报绝对误差		37.7		101.0		136.0	

种方法，其平均绝对误差值比另两种方法要小得多，且预报误差值比拟合误差值无明显增长趋势，预报效果比较稳定．按浙江省气象局现行评分办法，三年试报及实际预报平均成绩双重分析为90分，逐步回归63分，周期分析47分．

六、结论和讨论

(1)逐步回归双重分析是运用逐步回归技术同时进行前期因子筛选和提取预报对象的显著周期分量，并给予不同的权重系数，它既考虑到前期因子的支配作用，又兼顾了气象要素本身周期性变化的影响，这样有利于减少预报误差，提高预报精度．

(2)从目前所做的工作来看，逐步回归双重分析的预报效果明显优于周期分析和多元分析，预报效果也较为稳定．实践证明，该方法可以成为基层气象台站制作长期天气

预报的重要工具之一.

(3)在给定的F检验水平下，有时选入方程的变量全部都是前期因子，这正是逐步回归双重分析的独到之处. 当预报对象序列周期性变化不显著时，它能自动忽略周期项，无需进行人工干预.

(4)有些物理意义比较清楚的前期因子和周期分量，有时用逐步回归双重分析时不能入选，此时，也可仿照回归分析强制进入预报方程，这样有利于稳定方程的预报效果.

(5)在进行逐步回归双重分析时，预报对象序列应基本遵循正态分布，否则会影响预报效果，甚至导致失败.

参　考　文　献

[1] 李道松，1981，逐步回归与周期迭加相结合做长期预报的实验，中长期水文气象预报文集，第二集，水利电力出版社.

[2] 洋仰耕、余志忠，1983，用回归残差进行周期分析作月降水量预报，浙江气象科技，第1期.

[3] 魏凤英、赵濂、张先恭，1983，逐步回归周期分析，气象，第2期.

[4] 李邦宪，1986，试用逐步回归周期分析作月降水预报，浙江气象科技，第4期.

[5] 中国科学院数学研究所统计组编，1977，方差分析，科学出版社.

[6] 冯士雍，1985，回归分析方法，科学出版社.

[7] 王宗皓、李麦村等，1974，天气预报中的概率统计方法，科学出版社.

勘　误　表

本刊12卷4期中有如下差错，需进行勘误：

页	行	误	正
339	图1	锋生函数 $(Q,\nabla\theta)$	锋生函数 $(Q\cdot\nabla\theta)$
340	倒5	无辐散风、辐散风系统移动对…	无辐散风、辐散风，系统移动对…
340	倒9	$\zeta_a\nabla^2x-K\nabla\omega x\frac{\partial V}{\partial P}$	$\zeta_a\nabla^2x-K\cdot\nabla\omega x\frac{\partial V}{\partial P}$
387	7	X_1、X_4、…	X_1、X_3、X_4、…