

用CAR模型作年旱涝长期天气预报的研究¹⁾

郑洪初

(陕西安康地区气象局, 725000)

提 要

年旱涝时间序列是一动态系统。我们采用特殊的CARMA模型即带受控制项的自回归模型(CAR)对动态系统进行建模。并试图用较少的状态变量来描述系统的宏观行为。研究结果认为:用3个受控变量就可以描述安康市的年旱涝演变规律,并建立了CAR模型。用CAR模型作年旱涝预报能达到较高的精度,特别是大旱、大涝年预报效果最佳。试用表明预报与实况的相对误差为8.7%—12.6%。

关键词: 旱涝 长期预报 CAR模型

安排农业生产需要半年或一年以上的旱涝、冷暖长期天气预报,并根据旱涝和冷暖天气情况选择适宜的品种,确定适宜的播种期,避免或减轻旱涝和低温冷害对农作物产量和品质的影响,为确保稳产高产、优质、低耗提供决策依据。研究年旱涝预报模型,作好年旱涝预报具有重要的现实意义。

么枕生认为^[1-3],自回归模式用于统计预报最为适宜。他证明可用偏相关系数的 t 检验去确定自回归模式的阶。

1990年李翠华、么枕生^[4]用非线性自激励门限自回归对北京511年(1470—1980)历史旱涝记录变换的湿涝(干旱)游程记录进行了模拟和预报,解决了长期以来预报方程不能随转折点变更的问题。李邦宪^[5]针对经典多层递阶预报方法的某些不足,提出一种改进的多层递阶预报方法——多层递阶周期分析。他用时间序列的显著周期分量取代模型中的自回归部分,使其更合理地反映气象要素自身演变规律。并建立了逐步回归双重分析预报模型。但未给出阶数的确定方法。我们要研究的旱涝长期天气预报模型是带受控

制项的自回归滑动平均模型,简记为CARMA。由于这个模型的建模十分复杂,邓自立、郭一新等人^[6]提出用CAR模型代替CARMA对动态系统进行统一建模的新观点。两者之间的关系是任何的CARMA模型均可用充分高阶的CAR模型来逼近到任意精度,实际上则可用CAR模型对动态系统进行统一建模。下面将介绍如何建立CAR模型。

1 年旱涝CAR模型的建模和预报

年旱涝时间序列是一动态系统。动态系统可用带受控制项的自回归滑动平均CARMA模型来描述^[7]。它可用阶数 p, q 充分大的带受控制项的CAR模型所代替。

1.1 一个控制变量的CAR模型的建模

一个控制变量的CAR模型可写成如下形式:

$$y_t = a_1 y_{t-1} + \dots + a_n y_{t-n} + b_0 u_t + b_1 u_{t-1} + \dots + b_n u_{t-n} + \varepsilon_t \quad (1)$$

从式(1)可以看出,CAR模型由两部分组成,就时间序列 y_t 而言,一部分反映了时间序列本身的演变规律,另一部分则反映了

1) 参加者还有李敬云、屈善良、刘堂平、周书云等。

控制时间序列变化的系统对时间序列的影响程度。用CAR模型研究天气、气候的变化规律并用于中、长期和气候预报我们认为是适宜的。

自回归CAR模型是从低阶开始对系统进行建模,然后逐步增加模型的阶数,用 F 检验对这些模型进行自动筛选,采用带遗忘因子的最小二乘法可得到模型参数的一致估计,而且计算量小,计算简单,便于在气象台站推广应用。

令

$$\begin{cases} Q^T = (a_1 \cdots a_n \quad b_0 \cdots b_n) \\ Z_t^T = (y_{t-1} \cdots y_{t-n} \quad u_t \quad u_{t-1} \cdots u_{t-n}) \end{cases} \quad (2)$$

则式(1)写成最小二乘结构

$$y_t = Z_t^T Q + \varepsilon_t \quad (3)$$

其中 T 为转置号,根据式(1)在时刻 N , $\hat{Q}$ 的递推最小二乘估计值为:

$$\begin{cases} \hat{Q}_t = \hat{Q}_{t-1} + k_t (y_t - Z_t^T \hat{Q}_{t-1}) \\ k_t = \frac{P_{t-1} Z_t}{\alpha + Z_t^T P_{t-1} Z_t} \\ P_t = (I - k_t Z_t^T) P_{t-1} \end{cases} \quad (4)$$

为了避免在计算过程中 z_t 中的 y_{t-1} 和 u_t 出现负下标,不妨设递推时刻 $t=n$ 开始,可令初值为 $\hat{Q}_n = (0, \dots, 0)^T$, $p_n = \alpha I$ (α 是很大的正数,例如 $\alpha=10^4$, I 是单位矩阵)。

用 N 组观测数据所得到的CAR(n)模型的残差平方和 $S(n)$ 为

$$S(n) = \sum_{t=n+1}^N \varepsilon_t^2 \quad (5)$$

其中残差 $\hat{\varepsilon}_t$ 为

$$\begin{aligned} \hat{\varepsilon}_t &= y_t - \hat{a}_1(N)y_{t-1} - \\ &\cdots - \hat{a}_n(N)y_{t-n} - \hat{b}_0(N)u_t - \\ &\cdots - \hat{b}_n(N)u_{t-n} \end{aligned} \quad (6)$$

1.2 CAR模型阶数的选定(模式的鉴别)

CAR模型阶数的确定方法是基于对已知的 N 对观测数据($u_t, y_t, t=1, \dots, N$)由低阶到高阶逐步递增的对系统拟合CAR模型,并依次对相邻的两个CAR模型用 F 检验来确定模型阶数的适合性。

阶数的 F 检验如下:

构造统计量, H_0 :CAR(n)模型是适合的(即 $\hat{a}_{n+1}=0$,且 $\hat{b}_{n+1}=0$), H_1 :CAR(n)模型是

不适合的(即 $\hat{a}_{n+1}, \hat{b}_{n+1}$)至少有一个不为零。对于比较CAR(n)与CAR($n+1$)而言,统计量:

$$F = \frac{S(n) - S(n+1)}{S(n+1)} \cdot \frac{N - 2n - 3}{2} \quad (7)$$

其中 $S(n)$ 为低阶模型的残差平方和, $S(n+1)$ 为高阶模型的残差平方和。置信度 $\alpha=5\%$,令 $m=2, N=120$,查 F 分布表 $F_\alpha=3.07$,也可取 $m=2, N=\infty, F_\alpha=3$, m 和 N 为自由度。若 $F < F_\alpha$ 则CAR(n)是适合的(F 检验不显著); $F \geq F_\alpha$ 则是不适合的(F 检验显著)。

假设真实模型的阶是 n ,可由低阶开始,逐步增加的建立CAR(1),CAR(2),...CAR(n),CAR($n+1$),且令两模型间用 F 检验,直至找到适合的CAR(n)模型为止,一共需建立($n+1$)个CAR模型。

1.3 模式的估计(真实阶及时滞的 F 检验)

虽然得到了适合的CAR(n),但其中的某些参数 $\hat{a}_n, \hat{a}_{n-1}, \dots$,或者 $\hat{b}_n, \hat{b}_{n-1}, \dots$ 的估计值可能非常小,接近于零。这暗示自回归部分的阶或控制部分的阶比 n 小,实际上可能某些参数 $\hat{a}_n, \hat{a}_{n-1}, \dots$,或 $\hat{b}_n, \hat{b}_{n-1}, \dots$,为零。另一方面可能估算值 $\hat{b}_0 = \hat{b}_1 = \dots = \hat{b}_{n-1} = 0$ 即系统的时滞是 R 。

上述两种情况归结为,统计假设检验所得到的CAR(n)中的某些参数为零,从而最终决定系统的真实阶及时滞和相应的真实模型的参数估计。

由于白噪声 ε_t 的标准差的无偏估计为:

$$\hat{\sigma} = \sqrt{\frac{S(n)}{N - n}} \quad (8)$$

这是因为所有的观测数据不是 N 而是 $N-n$,而CAR(n)的未知参数个数是 $2n+1$ 。

在式(4)中 P_N 恰是参数 Q 的估值 $\hat{Q}_N$ 的误差协方差阵,即

$$P_N = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1,2n+1} \\ P_{12} & P_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ P_{1,2n+1} & \cdots & \cdots & P_{2n+1,2n+1} \end{bmatrix} \quad (9)$$

由正态分布性质 $\hat{Q}_N$ 的 $(1-\alpha)\%$ 的置信区间

简记为

$$\hat{Q}_i(N) \pm 1.96 \sqrt{P_{ii}} \delta (i = 1, 2, \dots, 2n+1) \quad (10)$$

由此可得到真实阶和时滞的 F 检验方法如下:

a. 判断哪些参数的估计值 $\hat{Q}_i(N)$ 的 95% 的置信区间包括零点。

b. 从所得到的 $CAR(n)$ 模型中删去这些参数后, 重新用带遗忘因子的递推最小二乘法建立较少参数的 $CAR(n)$ 模型。

c. 用 F 检验法对原来拟合的 $CAR(n)$ 和重新建立的较少参数的 $CAR(n)$ 模型进行检验, 如不显著, 则较少参数的 $CAR(n)$ 模型是真实模型; 如显著, 则原来的 $CAR(n)$ 模型为真实模型。

用 F 检验后, 较少参数的 $CAR(n)$ 为真实模型, 如所删去的参数为 $\hat{a}_{L+1}, \dots, \hat{a}_n$ 则自回归部分的阶是 L ; 如删去的参数为 $\hat{b}_{m+1}, \dots, \hat{b}_n$ 则控制部分的阶为 m ; 如删去控制部分的参数为 $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_{R-1}$ 则系统的时滞是 R 。

2 多个控制变量的 CAR 模型的建模

在实际系统中, 影响 y_t 的因素有很多个, 如果要建立由几个主要控制变量与 y_t 的关系, 而且要求求出时滞和阶数, 则可建立多输入单输出的 CAR 模型, 多输入单输出 CAR 模型的标准形式为:

$$\begin{aligned} y_t = & a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_n y_{t-n} + b_0 u_{1,t} + b_1 u_{1,t-1} + \dots \\ & + b_n u_{1,t-n} + c_0 u_{2,t} + c_1 u_{2,t-1} + \dots \\ & + c_n u_{2,t-n} + \dots + \varepsilon_t \end{aligned} \quad (11)$$

在编制程序时只需将(2)式的 $\hat{Q}_i, Z_i$ 变为:

$$\begin{aligned} \hat{Q}^T = & (a_1 \dots a_n b_0 \dots b_n c_0 \dots c_n \dots) \\ Z_i^T = & (y_{t-1} \dots y_{t-n}, u_{1,t}, \dots u_{1,t-n}, u_{2,t}, \dots u_{2,t-n} \dots) \end{aligned} \quad (12)$$

3 安康市年旱涝长期天气预报 CAR 模型的建模

我们采用中央气象台长期预报科提供的亚欧地区环流指数(45—65°N, 0—150°E)、亚洲地区环流指数(45—65°N, 60—150°E)和其它大气环流资料, 选取 3 个控制变量建立了安康市年旱涝预报模型。采用该模型可

在 12 月底以前作出第二年的旱、涝预报。3 个控制变量是:

u_1 : 前一年 500hPa 亚洲地区 5 月纬向环流指数; u_2 : 前一年 2 月 500hPa 亚洲地区经向环流指数; u_3 : 前一年 10 月 500hPa 亚洲地区经向环流指数。

按上述步骤, 从低阶开始逐步增加阶数, 建立 $CAR(1), CAR(2), \dots, CAR(n), CAR(n+1)$ 并用 F 检验直至找到适合的 $CAR(n)$ 模型为止。辨识结果如表 1。

表 1

参数	CAR(1)	CAR(2)	CAR(3)	真实的 CAR(2)
a_1	-0.2933	-0.5443	-0.6408	-0.5456
a_2		-0.8191	-0.9251	-0.8157
a_3			-0.2100	
b_0	-24.9086	-0.5679	-0.5246	
b_1	-7.7126	-40.4258	-37.1105	-40.5549
b_2		-33.8518	-41.8491	-34.3175
b_3			-1.8374	
c_0	85.6374	135.1430	137.4828	135.0154
c_1	50.8655	82.3742	110.4029	82.9968
c_2		65.1764	82.6344	64.9589
c_3			10.6011	
d_0	79.9581	59.7941	54.9020	59.2299
d_1	25.4570	81.9118	84.4082	82.5322
d_2		50.3975	67.5031	50.9942
d_3			-5.2097	
S_n	5168.6764	1844.2434	1751.0940	1846.0195
适合的阶	CAR(2)			

安康市年旱涝 CAR 预报模型为:

$$\begin{aligned} y_t = & -0.5456 y_{t-1} - 0.8157 y_{t-2} - \\ & 40.5549 u_{1,t-1} - 34.3175 u_{1,t-2} + \\ & 135.0154 u_{2,t} + 82.9968 u_{2,t-1} + \\ & 64.9589 u_{2,t-2} + 59.2299 u_{3,t} + \\ & 82.5322 u_{3,t-1} + 50.9942 u_{3,t-2} \end{aligned} \quad (13)$$

从式(13)可以看出: 安康市的旱涝不仅受自身演变规律的影响, 而且受亚洲地区 500hPa 5 月纬向环流指数和亚洲地区 500hPa 2 月、10 月经向环流指数的控制和影响。用该模型对安康市 1960—1989 年的旱涝指数进行了模拟计算, 预报值和实际值的平均相对误差 7.46%, 现列出大旱, 大涝年的模拟结果如表 2。

(下转封二)

表 2 用 CAR 模型模拟安康市大旱、大涝年降水量/mm

年份	1962	1965	1966	1972	1976	1984	1974	1979	1983	1984	1987
真实值	60.9	67.9	54.0	64.5	63.3	101.7	104.7	100.1	101.9	103.5	102.1
预报值	65.3	75.2	64.8	62.5	68.2	87.1	98.7	94.8	92.3	97.6	97.7
相对误差	-7.37	-10.81	-20.07	3.14	-7.89	14.31	5.67	5.29	9.37	5.69	4.26

1990、1991 年安康市年旱涝指数试报情况,按式(13)得:

$$y_{90} = 87.6$$

$$\text{实况 } 80.6 \quad \text{相对误差 } 8.7\%$$

$$y_{91} = 68.9$$

$$\text{实况 } 78.8 \quad \text{相对误差 } 12.6\%$$

4 结语

4.1 用 CAR 模型作年旱涝长期预报能达到较高的精度,收到满意的效果。特别是大旱、大涝年的预报效果最佳。历史上(1960—1989 年)共有 6 年年降水量大于 1000mm,预报的平均相对误差只有 7.43%,5 年年降水量小于 700mm,预报的平均相对误差只有 9.85%。历史上的大旱、大涝年均能作出准确的预报。如 1974 年的年降水量 1047mm,预报 987mm;1984 年的年降水量 1035mm,预报 976mm。1972 年的年降水量 646mm,预报 625mm。

4.2 寻找影响和控制旱涝变化的因子还没有一套成功的方法。虽然我们用逐步回归方法筛选出控制和影响安康地区旱涝变化的 3 个因子,但与安康地区的旱涝指数不能建立很好的 CAR 预报模型。可能的原因是山区

台站的降水量资料的代表性较差,而安康市是国家基准站,资料的代表性较好。

4.3 在气候学研究和长期天气预报工作中,往往要用较长时间的资料,必须注意资料的均一性。建立的 CAR 模型所用的 500hPa 环流资料是 B 模式提供的,而今后用于制作预报的 500hPa 环流资料是 T_{12} 模式的,虽然李小泉认为“差别是不大的”,但也应引起注意,差别总是有的。

参考文献

- 1 么枕生. 自回归模式的拟合. 大气科学, 1982, 5(4): 578—585.
- 2 么枕生. 长期天气预报线性自回归模式的拟合. 气象科技, 1984(4).
- 3 么枕生. 一个用偏相关筛选建立多元回归方程的方法. 气象科技资料, 1977(5).
- 4 李翠华、么枕生. 应用自激励门限自回归模式对旱涝游程序列的模拟和预报. 气象学报, 1990, 48(1).
- 5 李邦亮. 多层递阶周期分析. 气象, 1988, 14(11).
- 6 邓自立、郭一新. 动态分析及其应用. 辽宁科学出版社, 1985.
- 7 郭一新. 系统工程 BASIC 程序设计. 北京: 科学技术出版社, 1987.
- 9 李小泉等. 国家气象中心新旧业务系统月平均环流资料的比较. 气象, 1992, 3.

The Research on Annual Drought and Long-term Flood Prediction by CAR Model

Zheng Hongchu

(Meteorological Service of Ankang Prefecture, Shanxi Province 725000)

Abstract

The time series of annual drought and flood in Ankang prefecture, Shanxi province is of a dynamic system. Using the controlled auto-regression model(CAR), a special CARMA model, the system is modelled. The CAR prediction model of three controlled variables has been obtained, and it is applied to predict annual drought/flood tendency in Ankang region. Its accuracy is very well.

Key Word: drought and flood long-term prediction CAR model