

一种新的时间序列预报方法

周家斌

(中国科学院大气物理研究所)

提 要

本文将车贝雪夫多项式应用于时间序列的研究,提出了一种新的时间序列预报方法。这种方法精度较高,简便易行,可以用计算机算,也可以手工计算。文中给出了预报公式,估计了预报误差,并给出了预报实例。

一、引 言

单站时间序列的研究,历来受到气象台站的重视。特别是在县气象站,本站气象要素的历史演变曲线更是一种基本的预报工具。

五十年代,杨鉴初^[1]提出了历史曲线法,为我国气象台站应用单站要素曲线做长期预报探索了一条新路。这一方法至今仍在台站广泛应用。但是,这一方法缺乏必要的数学处理,因而更多地依赖于经验。

六十年代以来,平稳时间序列方法被引入气象预报,近年来更发展成为 ARIMA 模型,并在实践中取得了较好的效果。但是,一般时间序列的计算方法往往比较繁复,难以在气象站普及。

本文将车贝雪夫多项式应用于时间序列的研究,提出一种新的时间序列预报方法。

二、预 报 公 式

设在 t 轴上有 T_0 个格点: $t = 1, 2, \dots, T_0$ 。各格点上的气象要素值分别记为 $Z(1)$, $Z(2), \dots, Z(T_0)$ 。现在需要预报 $T_0 + 1$ 时刻的值,即 $Z(T_0 + 1)$ 。

在格点 $t = 1, 2, \dots, T_0, T_0 + 1$ 上将 $Z(t)$ 用车贝雪夫多项式展开,即^[2]

$$\tilde{Z}_{K_0}(t) = \sum_{k=0}^{K_0} A_k \varphi_k(t) \quad (t = 1, 2, \dots, T_0 + 1) \quad (1)$$

特别有:

$$\tilde{Z}_{K_0}(T_0 + 1) = \sum_{k=0}^{K_0} A_k \varphi_k(T_0 + 1) \quad (2)$$

上二式中 $\tilde{Z}_{K_0}(t)$ 为格点 t 处 $Z(t)$ 的拟合值, $\varphi_k(t)$ 为 k 阶归一化车贝雪夫多项式在格点 t 处的值, K_0 为(1)、(2)式截止阶数, A_k 为车贝雪夫系数, 由下式求出^[2]:

$$A_k = \sum_{t=1}^{T_0+1} Z(t)\varphi_k(t) \quad (k=0, 1, 2, \dots, K_0) \quad (3)$$

上式指出, A_k 依赖于格点 T_0+1 处的函数值 $Z(T_0+1)$, 因而是未知的. 如果我们能用 $Z(1), Z(2), \dots, Z(T_0)$ 求出 A_k 的近似值 A'_k , 则可求出 $\tilde{Z}_{K_0}(T_0+1)$ 的近似值为:

$$\tilde{Z}'_{K_0}(T_0+1) = \sum_{k=0}^{K_0} A'_k \varphi_k(T_0+1) \quad (4)$$

下面我们就来求 A'_k .

设 $Z(T_0+1)$ 的初估值为 $Z^{(0)}_{K_0}(T_0+1)$, 用之代替(3)式中的 $Z(T_0+1)$, 得 A_k 的一级近似值 $A^{(1)}_k$, 即

$$A^{(1)}_k = \sum_{t=1}^{T_0} Z(t)\varphi_k(t) + Z^{(0)}_{K_0}(T_0+1)\varphi_k(T_0+1) \quad (5)$$

用 $A^{(1)}_k$ 代替(2)式中的 A_k , 则得 $Z(T_0+1)$ 的一级近似值:

$$Z^{(1)}_{K_0}(T_0+1) = \sum_{k=0}^{K_0} A^{(1)}_k \varphi_k(T_0+1) \quad (6)$$

同样, 以 $Z^{(1)}_{K_0}(T_0+1)$ 代替(3)式中的 $Z(T_0+1)$, 可求得 A_k 的二级近似值为:

$$A^{(2)}_k = \sum_{t=1}^{T_0} Z(t)\varphi_k(t) + Z^{(1)}_{K_0}(T_0+1)\varphi_k(T_0+1) \quad (7)$$

从而由(2)式可求得 $Z(T_0+1)$ 的二级近似值为:

$$Z^{(2)}_{K_0}(T_0+1) = \sum_{k=0}^{K_0} A^{(2)}_k \varphi_k(T_0+1) \quad (8)$$

一般地, 有:

$$A^{(n)}_k = \sum_{t=1}^{T_0} Z(t)\varphi_k(t) + Z^{(n-1)}_{K_0}(T_0+1)\varphi_k(T_0+1) \quad (9)$$

$$Z^{(n)}_{K_0}(T_0+1) = \sum_{k=0}^{K_0} A^{(n)}_k \varphi_k(T_0+1) \quad (10)$$

$Z^{(n)}_{K_0}(T_0+1)$ 就是 T_0+1 处 Z 的预报值.

三、误差估计

为简便计, 引进如下记号及定义:

$$Z = Z(T_0+1) \quad (11)$$

$$\tilde{Z} = \tilde{Z}_{K_0}(T_0+1) \quad (12)$$

$$Z^{(n)} = Z^{(n)}_{K_0}(T_0+1) \quad (13)$$

$$\varepsilon^{(v)} = \varepsilon_{K_0}^{(v)}(T_0 + 1) = Z^{(v)} - Z \quad (v = 0, 1, 2, \dots) \quad (14)$$

$$\tilde{\varepsilon} = \tilde{\varepsilon}_{K_0}(T_0 + 1) = \tilde{Z} - Z \quad (15)$$

$$\sigma = \sigma_{K_0}(T_0 + 1) = \sum_{k=0}^{K_0} \varphi_k^2(T_0 + 1) \quad (16)$$

由(14)、(6)、(2)、(5)、(3)、(15)、(16)式得

$$\varepsilon^{(1)} = Z^{(1)} - Z = (Z^{(1)} - \tilde{Z}) + (\tilde{Z} - Z) = \sigma \varepsilon^{(0)} + \tilde{\varepsilon} \quad (17)$$

同理,

$$\varepsilon^{(2)} = \sigma \varepsilon^{(1)} + \tilde{\varepsilon} \quad (18)$$

将(17)式代入得:

$$\varepsilon^{(2)} = \sigma^2 \varepsilon^{(0)} + (1 + \sigma) \tilde{\varepsilon} \quad (19)$$

用数学归纳法可证:

$$\varepsilon^{(v)} = \sigma^v \varepsilon^{(0)} + \sum_{r=0}^{v-1} \sigma^r \tilde{\varepsilon} \quad (20)$$

引进记号:

$$\varepsilon_1^{(v)} = \sigma^v \varepsilon^{(0)} \quad (21)$$

$$\varepsilon_2^{(v)} = \sum_{r=0}^{v-1} \sigma^r \tilde{\varepsilon} \quad (22)$$

则

$$\varepsilon^{(v)} = \varepsilon_1^{(v)} + \varepsilon_2^{(v)} \quad (23)$$

可见,误差 $\varepsilon^{(v)}$ 由 $\varepsilon_1^{(v)}$ 与 $\varepsilon_2^{(v)}$ 两部分组成,其中 $\varepsilon_1^{(v)}$ 与初估值的误差 $\varepsilon^{(0)}$ 有关, $\varepsilon_2^{(v)}$ 与拟合值 $\tilde{Z}$ 的误差 $\tilde{\varepsilon}$ 有关. 此处所谓拟合值 $\tilde{Z}$, 系指当 $Z(T_0 + 1)$ 已知时由(2)、(3)两式算出的 $\tilde{Z}$ 值. 当 $K_0 = T_0$ 时, $\tilde{Z} = Z$, $\tilde{\varepsilon} = 0$. 因此,拟合值误差 $\tilde{\varepsilon}$ 就是由于截止阶数 $K_0 < T_0$ 引起的误差.

可以证明:

$$\sigma_{K_0}(T_0 + 1) < 1 \quad (K_0 < T_0) \quad (24)$$

利用(24),由(21)–(23)式可得:

$$\lim_{v \rightarrow \infty} \varepsilon_1^{(v)} = 0 \quad (25)$$

$$\lim_{v \rightarrow \infty} \varepsilon_2^{(v)} = \frac{\tilde{\varepsilon}}{1 - \sigma} \quad (26)$$

$$\lim_{v \rightarrow \infty} \varepsilon^{(v)} = \frac{\tilde{\varepsilon}}{1 - \sigma} \quad (27)$$

由此可见,迭代是收敛的. 其中与初估值误差 $\varepsilon^{(0)}$ 有关的误差收敛到 0, 与拟合值误差 $\tilde{\varepsilon}$ 有关的误差 $\varepsilon_2^{(v)}$ 收敛到 $\frac{\tilde{\varepsilon}}{1 - \sigma}$. 由于 $0 < \sigma < 1$, 故 $\varepsilon_1^{(v)}$ 与 $\varepsilon^{(0)}$ 同号, $\varepsilon_2^{(v)}$ 与 $\tilde{\varepsilon}$ 同号, 亦即这两部分误差都随迭代次数 v 的增加而单调变化. 因此,当 $\varepsilon^{(0)}$ 与 $\tilde{\varepsilon}$ 同号时,两部分误差叠加;当 $\varepsilon^{(0)}$ 与 $\tilde{\varepsilon}$ 反号时,两部分误差可抵消一部分,从而得到较小的误差.

四、决定误差的几个因子

1. 初估值误差 $\epsilon^{(0)}$

各种情况下 $\epsilon_1^{(v)}$ 、 $\epsilon_2^{(v)}$ 的变化可由 (21)–(23), (25)–(27) 式得出.

根据分析, 如果初估值满足如下两个条件, 则迭代至某一临界次数 ν_c , 其预报误差最小. 即:

A. $\epsilon^{(0)}$ 与 $\bar{\epsilon}$ 反号,

$$B. |\epsilon^{(0)}| > \frac{|\bar{\epsilon}|}{\sigma}.$$

这一情况见图 1.

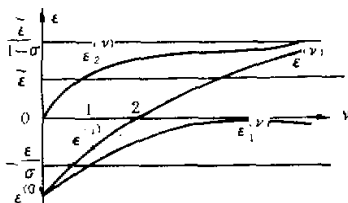


图 1 理想初估值情况下迭代误差的变化

2. 截止阶数 K_0

当 K_0 增大时有两种相反的影响, 即: 1) K_0 增大, $\sigma_{K_0}(T_0 + 1)$ 增大, 因而 $\epsilon_1^{(v)}$ 增大.

2) K_0 增大, $\bar{\epsilon}$ 减小, 但 $\sum_{r=0}^{v-1} \sigma^r$ 增加. 因而 $\epsilon_2^{(v)}$ 在某个中间的 K_0 值时最小. 因此, 总的说来, 在某个适中的 K_0 处 $\epsilon^{(v)}$ 最小.

3. 拟合误差 $\bar{\epsilon}$

$\bar{\epsilon}$ 影响 $\epsilon_1^{(v)}$ 及 $\epsilon^{(m)}$, 因此求得较小的 $\bar{\epsilon}$ 对改进预报相当重要. 提高 K_0 , 可减小 $\bar{\epsilon}$, 但大的 K_0 有前述弊病. 因此, 为了减小 $\bar{\epsilon}$, 应对序列作预处理, 使其变得比较光滑.

4. 迭代次数 ν

ν 与 $\epsilon^{(v)}$ 的关系由 (20) 式给出. 即 ν 愈大, 初估值误差的影响愈小, 拟合值误差的影响愈大.

5. 格点数 $T_0 + 1$

T_0 不同时, $\sigma_{K_0}(T_0 + 1)$ 不同, $\bar{\epsilon}_{K_0}(T_0 + 1)$ 亦不同. 故 T_0 的选择对预报精度亦有影响.

五、多时刻的预报

前面讨论的都是 $T_0 + 1$ 这一时刻的预报, 现在再简单讨论一下多时刻的预报问题. 多时刻预报可通过两种途径实现.

1. 先做出 $T_0 + 1, T_0 + 2, \dots, T_0 + n - 1$ 时刻的预报, 令其为已知, 然后做出 $T_0 + n$ 时刻的预报. 公式是

$$Z_{K_0}^{(0)}(T_0 + n) = \sum_{k=0}^{K_0} A_k^{(0)} \varphi_k(T_0 + n) \quad (28)$$

其中

$$A_k^{(0)} = \sum_{t=1}^{T_0} Z(t) \varphi_k(t) + \sum_{t=T_0+1}^{T_0+n-1} Z_{K_0}^{(0)}(t) \varphi_k(t) + Z_{K_0}^{(0-1)}(T_0 + n) \varphi_k(T_0 + n) \quad (29)$$

2. 用 $t = 1, 2, \dots, T_0$ 时刻的资料, 同时作出 $T_0 + 1, T_0 + 2, \dots, T_0 + n$ 时刻的预报. 公式是:

$$Z_{K_0}^{(0)}(T_0 + r) = \sum_{k=0}^{K_0} A_k^{(0)} \varphi_k(T_0 + r) \quad (30)$$

其中

$$A_k^{(0)} = \sum_{t=1}^{T_0} Z(t) \varphi_k(t) + \sum_{t=T_0+1}^{T_0+n} Z_{K_0}^{(0-1)} \varphi_k(t) \quad (r = 1, 2, \dots, n). \quad (31)$$

对于(28)与(30)式的误差可做类似讨论.

由于本方法对时间序列的采样没有限制, 因而多时刻的预报也可用前述做一个时刻预报的方法作出. 例如, 用一年的采样间隔做次年的预报, 用两年的间隔做第三年的预报.

六、操作步骤

1. 历史资料的处理

A. 资料预处理

可针对不同的预报对象采取不同的预处理方法. 例如, 对降水量, 可首先求出距平百分率, 以减少季节影响, 然后再用下式标准化:

$$Z' = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}} \quad (32)$$

此时, 有

$$0 \leq Z' \leq 1 \quad (33)$$

B. 选择格点数 $T_0 + 1$.

- C. 对不同的 K_0 拟合, 求出 ε_{K_0} .
- D. 对不同的 K_0 用各种初估值进行迭代运算.
- E. 确定所需初估值的类型, 即从各种初估值中选出满足如下两个条件的初估值. a . $\varepsilon^{(0)}$ 与 ε 反号, b . $|\varepsilon^{(0)}| > \frac{|\varepsilon|}{\sigma}$.
- F. 以最常见的 ν_c 为所有个例的统一迭代次数, 由此得出历史资料的预报结果.
- G. 若迭代终值误差较小, 亦可用其作为预报结果. 这样可避免初估值类型的选择与 ν_c 之确定.
- H. 对不同 T_0 与 K_0 情况下的计算结果进行预报集成^[1], 求出最后的预报结果.

2. 实际预报

- A. 资料预处理: 同前.
- B. 选择格点数: 同前.
- C. 对不同的 K_0 用各种初估值进行迭代运算. 同前.
- D. 由迭代结果的变化情况决定所需初估值的类型. 以处理历史资料时确定的常见 ν_c 为预报用 ν_c , 并决定预报值.
- E. 若历史资料之迭代终值可用, 则亦可用实际预报时的迭代终值作为预报结果.
- F. 预报集成.

七、应用实例

应用上述方案, 对新安江地区年降水趋势进行了预报试验. 预报对象为新安江流域十个水文站一至九月降水总量的加权平均值. 十个测站是: 屯溪、渔梁、三阳坑、街口、上溪口、黟县、黄山、百罗畈、姜家、排岭. 资料年代为 1932—1981. 所有资料均进行了距平百分率标准化处理.

所取格点数有 $T_0 + 1 = 7, 9, 11$ 三种. 对不同 K_0 进行迭代, 初估值取为 $-4, -3.5, -3, \dots, -0.5, 0, 0.1, 0.2, \dots, 1, 1.5, 2.0, 2.5, \dots, 5$. 根据分析, $Z' \leq -1$ 与 $Z' \geq 2$ 两类情况必有一类为理想初估值. 因此, 在实际预报时, 只要判断上述两类初估值中何者属于此类即可. 以 1938 年为例, 当 $T_0 = 7, K_0 = 2$ 时, 迭代终值为 $Z'^{(00)} = Z' + \frac{\varepsilon}{1 - \sigma} = 0.861$. 如前所述, 该类初估值之 $\varepsilon^{(0)}$ 与 ε 反号, 而 $\varepsilon^{(00)}$ 与 ε 同号. 故此时真值必在 $\varepsilon^{(0)}$ 与 $\varepsilon^{(00)}$ 之间. 因此, 若 $Z'^{(00)} \geq 2$ 为理想初估值, 则真值 Z 必在 0.861 与 2 之间. 本例 $\sigma = 0.76$, 故终值误差为 $\varepsilon/0.24$. 此值一般较大. 因此, 若真值在 0.861 与 2 之间, 则其值甚大, 甚至超过 1, 而这是不可能的. 故判定 $Z^{(0)} \leq -1$ 为理想初估值. 由此看来, 将资料处理成 Z' , 不仅使序列比较光滑, 而且有利于初估值类型的决定.

图 2 给出 $T_0 + 1 = 11, K_0 = 1, \nu = 2, Z^{(0)} = -2$ 或 3^0 的计算结果作为例子. 其

1) 如前所述, $Z^{(0)} = -2$ 或 3 中必有一种属于理想初估值, 图 2 即为该类初估值的结果.

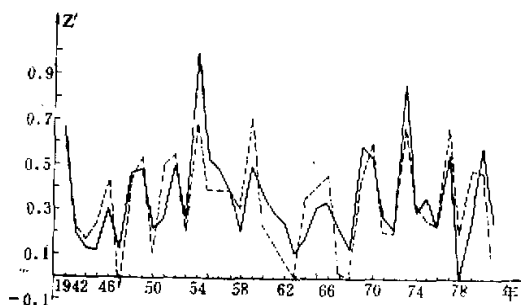


图2 新安江流域年降水趋势计算结果

中实线为实测值,虚线为计算值。易见,计算结果是比较好的,主要的旱涝年都拟合出来了。

尚需指出,上述结果尚未进行预报集成。

由(32)式可得:

$$Z = Z_{\min} + Z'(Z_{\max} - Z_{\min}) \quad (34)$$

对本例,有:

$$Z = 1.01Z' - 0.369 \approx Z' - 0.369 \quad (35)$$

因此, Z' 值与 Z 基本上只差一个常数因子,故图2基本上表示了距平百分率的预报情况。

令(35)式中 $Z = 0$, 得 $Z' = 0.365$, 因此, $Z' = 0.365$ 处相当于距平零线。

本文所介绍的方法于1982年由新安江电厂气象组用于业务预报。1983年的年降水趋势预报已由作者提供新安江电厂试用*, 全国一些水文气象台站也正在试用这一方法。

八、结 语

本文提出的时间序列预报方法比较简便,可以用计算机算,也可以手算。因而,这一方法除可用计算机大量试验外,尚可供没有计算机条件的水文气象台站使用。

本方法对时间序列的采样没有限制,因而长中短期预报均可使用。

参 考 文 献

- [1] 杨懿初,运用气象要素历史演变的规律性作一年以上的长期预报,气象学报,第24卷100—117,1953.
- [2] Багров Н. А., Аналитическое представление полей, Труды ЦИП, 64, 3—25, 1958.
- [3] 周家斌,关于预报集成问题,单站统计天气预报方法的研究,中国科学院大气物理研究所集刊,第3号,科学出版社,62—71,1975.

* 附记 作者用这一方法于1983和1984年向新安江水力发电厂作了1—9月降水总量预报。1983年预报值为1874毫米,实况为2027毫米(平均值为1486毫米),1984年预报值为1445毫米,实况值为1438毫米。

A NEW TYPE OF TIME-SERIES-FORECASTING METHOD

Zhou Jiabin

(Institute of Atmospheric Physics, Academia Sinica)

Abstract

By giving function value $Z(t)$ at grids $t = 1, 2, \dots, T_0$, the value of $Z(T_0 + 1)$ may be evaluated from the following Exps.

$$Z_{K_0}^{(0)}(T_0 + 1) = \sum_{k=0}^{K_0} A_k^{(0)} \varphi_k(T_0 + 1) \quad (1)$$

$$A_k^{(\nu)} = \sum_{t=1}^{T_0} Z(t) \varphi_k(t) + Z^{(\nu-1)}(T_0 + 1) \varphi_k(T_0 + 1) \quad (\nu = 1, 2, 3, \dots; K_0 < T_0) \quad (2)$$

where $Z_{K_0}^{(0)}(T_0 + 1)$ denotes the ν -th approximation to $Z(T_0 + 1)$, $Z^{(0)}(T_0 + 1)$ is the initial guess, ν , the iteration time, $\varphi_k(T_0 + 1)$, the value of normalized Chebyshev polynomial with k order at grid $T_0 + 1$, and $A_k^{(0)}$, the ν -th approximation to Chebyshev coefficient A_k .

The error of Exp. (1) in forecasting may be estimated with the following Exp.

$$\varepsilon^{(\nu)} = \varepsilon_1^{(\nu)} + \varepsilon_2^{(\nu)} = \sigma^\nu \varepsilon^{(0)} + \sum_{r=0}^{\nu-1} \sigma^r \tilde{\varepsilon} \quad (3)$$

where

$$\varepsilon^{(\nu)} = Z_{K_0}^{(\nu)}(T_0 + 1) - Z(T_0 + 1) \quad (4)$$

$$\varepsilon^{(0)} = Z^{(0)}(T_0 + 1) - Z(T_0 + 1) \quad (5)$$

$$\tilde{\varepsilon} = \tilde{Z}_{K_0}(T_0 + 1) - Z(T_0 + 1) \quad (6)$$

$$Z_{K_0}(T_0 + 1) = \sum_{k=0}^{K_0} A_k \varphi_k(T_0 + 1) \quad (7)$$

$$A_k = \sum_{t=1}^{T_0+1} Z(t) \varphi_k(t) \quad (8)$$

$$\sigma = \sum_{k=0}^{K_0} \varphi_k^2(T_0 + 1) < 1. \quad (9)$$

Here $\varepsilon^{(0)}$ is the error of initial guess and $\tilde{\varepsilon}$ is the fitting error, which are evaluated from Exps. (6)–(8).

There exist the relationships

$$\lim_{\nu \rightarrow \infty} \varepsilon_1^{(\nu)} = 0 \quad (10)$$

$$\lim_{\nu \rightarrow \infty} \varepsilon_2^{(\nu)} = \frac{\tilde{\varepsilon}}{1 - \sigma} \quad (11)$$

$$\lim_{\nu \rightarrow \infty} \varepsilon^{(\nu)} = \frac{\tilde{\varepsilon}}{1 - \sigma} \quad (12)$$

therefore, the Exp. (1) is convergent.

It is worth pointing out that $\varepsilon^{(\nu_e)} \approx 0$ for some ν_e when the sign of $\varepsilon^{(0)}$ is opposite to that of $\tilde{\varepsilon}$ and $|\varepsilon^{(0)}| > \frac{|\tilde{\varepsilon}|}{\sigma}$. In these conditions better forecast may be obtained.

The method proposed in this paper has been used in the forecast of precipitation over the Xin'an River Valley and better results have been obtained.