

不同数学模型在降水量资料序列 订正中适用性探讨*

宋超辉

(国家气象中心, 北京 100081)

提 要

采用逐步多元线性回归模型、一元线性回归模型、综合法模型和比值法模型对分布于全国 31 个台站的降水量资料进行了模型订正效果的试验与分析. 结果表明: (1) 逐步多元线性回归模型对年降水量序列的订正效果较好, 其相对拟合误差总平均在 0.08 以下, 而其它 3 种模型与月降水量序列的订正效果较差, 相对拟合误差总平均在 0.11 以上; (2) 若将年降水量序列相对拟合误差控制在 0.10 以内, 则要求确定订正方程式的平行资料年数为 10 年或以上, 要求订正站与基本站序列的相关系数在 0.85 以上; (3) 较湿润地区拟合误差较小, 反之较大.

关键词: 数学模型 降水量 相对拟合误差 订正效果

引 言

70 年代以来, 气候异常, 各地旱涝灾害频繁, 气候工作者更加重视对于降水量变化规律的研究, 但由于站址迁移、仪器变更等人为因素造成降水量资料的非均一性和不完整性, 从而影响了对其正确的使用. 因此, 对非均一性的资料进行订正, 对中断缺记录的资料插补, 对不够长的短资料序列延长(以下订正、插补、延长统称为序列订正)是十分必要的.

至今已有的序列订正方法研究成果^[1~5]中, 对降水量序列的研究较少, 尤其是投入到大批量资料的处理经验也不多. 因此, 对降水量资料序列订正方法及其适用性进行研究也是十分重要的.

1 资料与处理

本次试验用的站点由分布于全国的 31 个站点作为订正站, 并在每个订正站的周围邻近站中选取距离相近、相关性较好且具有较长平行资料的 2~5 个站作为基本站或基

* 本文系“九五”期间 96-908-04-08-1 课题资助.
1996-11-11 收到, 1997-01-19 收到修改稿.

本站的备选站(以下统称基本站)共 125 个,资料年代为 1957~1986 年,序列为 1、4、7、10 月和年的降水总量。

试验中要求基本站降水量资料都是均一的气候序列,为此,除满足上述条件外,均采用多元两段回归法^[6~8]对序列进行非均一性检验,将具有非均一性的站点剔除。

2 数学模型及方法

在相邻测站间气候相似性的原则指导下,根据相邻测站资料年际变化的相关性,用一定的样本资料进行拟合得到订正方程式,然后在这些方程式中代入基本站资料来计算订正站的资料。

2.1 一元线性回归模型

$$y_a = \beta_0 + \beta_1 x_a + \epsilon_a \quad (a = 1, 2, \dots, N) \quad (1)$$

其中 $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ 分别表示随机因素对订正站序列 y_a 的影响总和(下同); N 为样本容量,即表示 N 组观测数据; x_a 代表基本站序列;参数 β_0, β_1 的最小二乘法估算值分别为 b_0 和 b_1 ,于是可以得到一元线性回归方程。

$$\hat{y} = b_0 + b_1 x \quad (2)$$

2.2 逐步多元线性回归模型

$$y_a = \beta_0 + \beta_1 x_{a1} + \beta_2 x_{a2} + \dots + \beta_p x_{ap} + \epsilon_a \quad (a = 1, 2, \dots, N) \quad (3)$$

自变量 x_j (其中 j 为自变量的序号, $j=1, 2, \dots, p$)的偏回归平方和为:

$$\theta_j = b_j^2 / C_{jj} \quad (4)$$

其中, N 为样本容量,即表示 N 组观测数据; p 为基本站的个数; y_a 为订正站的序列; $x_{a1}, x_{a2}, \dots, x_{ap}$ 为基本站序列; \hat{y}_a 为所求的回归方程 $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$ 的第 a 个回归值; C_{jj} 为相关矩阵 C 对角线上的元素; b_0, b_1, \dots, b_p 分别为参数 $\beta_0, \beta_1, \dots, \beta_p$ 的最小二乘估计值。

2.3 综合法模型和比值法模型

这两个数学模型与一元线性回归的数学模型类似,只是为计算简便在一元线性回归方程的基础上作某些假设,通常一元线性回归方程为:

$$\hat{y}_a = \bar{y}_N + r(x, y) (\sigma_y / \sigma_x) (x_a - \bar{x}_N) \quad (5)$$

当假设 $r(x, y) = 1, (\sigma_y / \sigma_x) = k$ 时,得综合法模型

$$\hat{y}_a = \bar{y}_N + k(x_a - \bar{x}_N) \quad (6)$$

当假设 $r(x, y) = 1, (\sigma_y / \sigma_x) = (\bar{y}_N / \bar{x}_N) = k$ 时,得比值法模型

$$\hat{y} = k \bar{x}_a \quad (7)$$

上式中 $\bar{y}_N, \bar{x}_N, \sigma_y, \sigma_x$ 和 $r(x, y)$ 分别为样本容量为 N 的订正站与基本站的平均值、均方差和相关系数。

上述 4 种模型都是建立在回归分析理论基础之上,所以必须进行回归方程显著性和订正方程式适用性检验^[9],经检验确认置信水平($\alpha=0.10$)是有效的情况下,再分析经订正后的值 \hat{y}_a (回归估计值)与相应观测值 y_a 的差值(即拟合误差),拟合误差越小,效果

也就越好,但对于降水量资料而言,其绝对值随时间、空间变化很大,拟合误差的比较性差.为便于比较,本文一律采用相对拟合误差,即 $|\hat{y}_a - y_a| / \bar{y}_N$.

3 试验结果的分析

3.1 不同台站拟合误差的分析

为了比较不同台站的序列订正效果,对31个订正站的年降水量采用逐步多元回归模型进行拟合,并计算了拟合误差等有关值(表略).结果表明,多年平均降水量越少时,拟合误差越大,如喀什站;反之越小,如龙岩站.还发现,当多年平均降水量 <300 mm时,平均相对拟合误差 ≥ 0.11 ,当多年平均降水量 >300 mm时,平均相对拟合误差为 ≤ 0.10 ,当 ≥ 1000 mm时,平均相对拟合误差和最大相对拟合误差都有偏小的趋势,分别为 $0.05\sim 0.08$ 和 $0.12\sim 0.20$.

3.2 不同模型拟合误差的比较

为了寻找降水量序列订正的最佳模型,即拟合误差最小,试验采用以上4种模型对1957~1986年年降水量资料在25年以上,多年平均降水量在300 mm以上的27个站进行拟合,并统计了不同模型下的有关值(表1).从表1看到,4种模型的相对拟合误差,以逐步多元线性回归最小,总平均为0.079,其次是一元线性回归,比值法与综合法相当.很显然逐步多元回归比一元回归考虑的因子(基本站)多,反映出订正站与基本站之间的定量关系更加准确,其复相关系数从 $0.38\sim 0.90$ 提高到 $0.73\sim 0.97$,拟合误差最小,而其余两个方法因作了一些假设,这些假设与实际观测值毕竟是有差异的,所以拟合效果差一些.从表1中还可以看出,除逐步多元回归以外,3种模型的拟合误差虽稍有差别,但都超过0.10,误差较大.

表1 不同模型订正效果的比较

	拟合模型			
	比值法	综合法	一元回归	逐步多元回归
相关系数	0.38~0.90	0.38~0.90	0.38~0.90	0.73~0.97*
相对标准差	0.14~0.36	0.14~0.36	0.14~0.36	0.14~0.36
相对拟合误差总平均	0.120	0.121	0.111	0.079
最大相对拟合误差	0.76	0.69	0.61	0.54
最大相对拟合误差 ≤ 0.20 站数	1	2	1	10
最大相对拟合误差 ≤ 0.30 站数	7	6	8	17

注: * 为复相关系数

3.3 不同相关系数下的相对拟合误差

本试验采用一元线性回归,用多年平均降水量在300 mm以上各试验站资料进行了拟合效果统计(表2),结果表明订正站与基本站资料序列的相关系数越大,拟合误差越小,若将其控制在0.10以内,则要求相关系数 ≥ 0.85 .

表2 不同相关系数的相对拟合误差值

相关系数	样本数	平均相对拟合误差
<0.60	108	0.123
0.60~0.69	170	0.116
0.70~0.79	115	0.115
0.80~0.84	229	0.109
0.85~0.90	132	0.098
>0.90	无	无

3.4 不同平行资料年数拟合误差值的比较

建立订正方程式需要一定的样本容量,即平行资料的年数.本试验以逐步多元线性回归对年平均降水量>300 mm的试验站分别建立了平行资料年数为5、10、15、20、25年情况的多元线性回归订正方程式,然后用这些方程式计算所有 $x_{a1}, x_{a2}, \dots, x_{ap}$ 下的拟合值 \hat{y}_a 与实际观测值 y_a 的相对拟合误差值(表3).

表3表明,以平行资料长的,即大样本为基础建立的回归方程估计,其效果要好.对于小样本,如平行资料5年,即使相关系数较高(0.891~0.998),其拟合误差依然很大,反之样本年数较长(如25年),相关系数虽相对低一些(0.675~0.968),但订正效果较好,若将相对拟合误差控制在0.10以内,则要求平行资料10年以上.

表3 不同平行资料年数多站平均相对拟合误差值

	5年	10年	15年	20年	25年
复相关系数	0.891~0.998	0.595~0.989	0.760~0.983	0.631~0.968	0.675~0.968
相对标准差	0.118~0.304	0.139~0.290	0.133~0.254	0.138~0.256	0.141~0.293
站平均相对拟合误差	0.062~0.243	0.054~0.235	0.056~0.199	0.044~0.117	0.044~0.110
相对拟合误差总平均	0.133	0.104	0.101	0.085	0.079
最大相对拟合误差	0.270~0.879	0.174~0.776	0.131~0.395	0.100~0.422	0.109~0.603

3.5 不同时段序列拟合效果的分析

本试验采用样本年数为25年以上且多年平均降水量>300 mm共27站平行资料的1、4、7、10月和年序列,采用逐步多元线性回归拟合进行了统计(表4).表4表明,月序列中无论哪一个月份,其相对拟合误差的总平均都超过0.25,但年序列的拟合误差的总平均在0.08以下,且站平均相对拟合误差也在0.10以内,所以年降水量序列是可以进行订正的.

表4 不同时段序列的相对拟合误差值

	1月	4月	7月	10月	年
复相关系数	0.63~1.00	0.47~0.99	0.60~0.96	0.64~0.99	0.73~0.93
复相关系数>0.90的站数	23	22	7	20	12
相对拟合误差的总平均	0.604	0.329	0.270	0.663	0.079
站平均相对拟合误差	0.11~2.01	0.07~0.84	0.11~0.56	0.10~4.09	0.04~0.10
站平均相对拟合误差>0.10站数	27	26	27	20	0
最大相对拟合误差	0.43~36.04	0.23~9.36	0.33~3.94	0.53~81.22	0.10~0.54
相对标准差	0.47~1.86	0.33~1.51	0.30~0.73	0.50~1.33	0.14~0.36

4 结 论

(1) 4种模型中,以逐步多元回归模型拟合误差最小,相对拟合误差为0.079,其它3种模型虽稍有区别,但误差较大,相对拟合误差均超过0.10.

(2) 要求建立订正方程式的资料是均一的,而且还要求一定的平行资料长度,资料越长,效果越好.若将相对拟合误差控制在0.10以内,则要求平行资料年数为10年或以上.

(3) 拟合误差大小与订正站和基本站的相关程度有密切关系,在满足一定的平行资料长度前提下,相关系数越高,拟合误差越小.若将相对拟合误差控制在0.10以内,则要求相关系数在0.85以上.

(4) 不同时段降水量序列订正的拟合误差也不相同,年序列较小,相对拟合误差在0.08以下;月序列较大,均在0.25以上.

(5) 若以多年平均降水量的多少表示湿润状况,则较湿润地区效果较好,拟合误差较小.试验表明若将平均相对拟合误差控制在0.10以内,则要求多年平均降水量达到300 mm以上.

参 考 文 献

- 1 宋超辉,孙安健.非均一性气温序列订正方法的研究.高原气象,1995,14(2):215~219.
- 2 屠其璞.平均气温序列延长订正的讨论.南京气象学院学报,1979,2(2):193~200.
- 3 屠其璞.一种气温场序列的延长插补方法.南京气象学院学报,1986,9(1):19~30.
- 4 陈万隆,翁笃鸣,钱林清.中区域山区温度和降水推算若干问题的讨论.南京气象学院学报,1980,3(1):95~104.
- 5 Thamas R Kerl, Claude N and Williams J R. An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.*, 1982, 12: 1744~1763.
- 6 Andrew R Solow. Testing for climate change: An application of the two-phase regression model. *J. Climate Appl. Meteor.*, 1987, 26: 1401~1606.
- 7 Lucie Vincent G. Time series analysis; testing the homogeneity of monthly temperature series. Atmospheric Environment Service Canadian Climate Centre and York University, July 1990, 1~49.
- 8 Gullett D M, Vincent L and Malone L H. Application of multiple-phase models with mathematical change point. A Publication of the Canadian Climate Centre, 1991. 1~47.
- 9 么枕生,丁欲国.气象统计.北京:气象出版社,1990. 775~820.

RESEARCH ON ADJUSTING EFFECT FOR DIFFERENT MATHEMATICAL MODEL OF PRECIPITATION SERIES

Song Chaohui

(National Meteorological Center, Beijing 100081)

Abstract

Based on the models of successive multiple linear regression, one-variate linear regression, ratio value method and aggregate method the analyses and tests of precipitation series correction are carried out by using the data observed at thirty-one stations in twenty-five provinces of China. The results show that

(1) the correction effect of multiple linear regression method for annual precipitation data series is better than those of other methods, the average relative fitting error is less than 0.08, and the other is more than 0.11 for monthly data series.

(2) if the relative fitting error is less than 0.10, the time of paralld data requested in the equation should be longer than 10 years, the correlation coefficient above 0.85.

(3) if the precipitation is above 300mm, the error is less than 0.10 and if below 300 mm, the error is more than 0.11.

Key words: Mathematical model Precipitation Relative fitting error Correction effect