

夏季降水量与气温资料的恢复试验*

黄嘉佑

(北京大学物理学院大气科学系,北京 100871)

刘小宁 李庆祥

(国家气象中心气象资料室,北京 100081)

摘 要

该文对我国汛期降水量和气温缺测资料及资料均一性恢复的方法进行试验研究。对单个气候地区(东北区),以及全国 6 个气候地区用逐步回归和逐步判别方法进行缺测 1~5 年恢复试验。对东北地区内所有站点的试验表明,逐步判别分析方法比逐步回归分析方法恢复效果好些;选择邻站用判别分析方法建立的恢复模型,效果更好。全国气候地区试验表明,选择邻站的判别分析方法有较好的恢复效果。

关键词:资料恢复 均一性 判别分析 回归分析

引 言

在气候变化的研究中,气候资料的可靠性和精确性是气候变化研究的基础。由于早期观测仪器或者观测台站的变迁等等原因,使早期观测资料的均一性存在很多问题^[1]。另外,近年来年代际的气候变化研究十分活跃^[2~5],在我国某些台站有 100 年以上的降水量和气温资料,但是由于解放初期,在一些年份或月份有些测站缺少观测资料。因此较准确估计并恢复测站缺测的资料,对于我国年代际气候变化研究具有十分重要的意义。国内外对资料的均一性有不少研究。例如, Potter^[6]提出两变量正态分布检验对资料序列均一性进行检验。恢复的有关方法也有不少研究。Easterling 和 Peterson^[7]提出用回归分析与非参数统计量结合进行序列均一性进行检验。但是,缺测资料恢复的方法研究还不多。最近, Young 使用 3 种不同的方法,多元判别、多元回归和正态比方法,分别对美国 Arizona 和 Mexico 州的 111 个测站的月降水量进行插值试验^[8],对 19 世纪 90 个资料和 20 世纪的 260 个资料进行插值试验,用 3 个方法的预测值的中值作为 3 种方法的预测值。结果表明,多元判别有较好的插值效果。

对于我国的资料,特别是降水量和气温的夏季汛期季节资料用什么方法恢复比较合适?是否使用判别分析方法有好的效果?而判别分析方法涉及到首先对恢复对象资料进行分类,怎样分类?哪种分类方法?分多少类?怎样与回归分析方法进行比较?用多元

* 本文由国家科技基础性工作专项基金项目 2001 DE A30029-01 及国家气象中心课题“台站环境改变对我国气温观测的影响研究”(课题号为 ZK2002B-19)的资助。

2003-01-02 收到,2003-05-06 收到修改稿。

判别(回归)还是逐步判别(回归)?由于逐步判别和逐步回归都对进入方程的因子进行引入或剔除,用什么标准进行?如果用多元判别和多元回归,又怎样选择必须进入方程的因子?本文就是试图回答这方面问题,并进行比较试验研究。

在我国气候变化的研究中,最重要的研究对象莫过于夏季降水量和气温。因此本文研究缺测资料的恢复,主要是6~8月的降水量和气温,使用我国160测站资料(由中国气象局国家气象中心提供)。

1 资料与方法

我们选取逐步判别与逐步回归^[9]作为资料恢复试验的基本方法。

由于季节降水量和气温是气候要素,试验时先对我国降水量进行气候区划,气候区划分区方法见文献^[10],即把我国气候降水特点分为东北、华北、黄河流域、长江流域、华南和新疆等6个地区,各区测站数分别包含22,26,34,37,29和12个测站。

选取1951~2000年东北地区6~8月季节降水量和气温的22个站的数据。比较方法是把其中某站或某些站的数据看成为缺测,用整个地区的其他测站数据进行恢复。以比较哪个方法有较好的恢复效果。恢复值与实测值的差值和它们试验样本的均方误(平均平方根误差)是比较效果的评判标准。又考虑到气温与降水量的恢复值单位不同,而且即使是同一要素,由于站点地理位置不同,其气候变化幅度(即标准差)不同,为了消除要素和站点位置差异,也便于不同要素相互比较,引入标准误差,如下式所示,

$$e = \frac{\hat{x} - x_0}{s} \quad (1)$$

式中 \hat{x} 为要素恢复值, x_0 为要素观测值, s 为要素多年的样本标准差。对此标准误差再计算它们试验样本的均方误,称为标准均方误,即

$$es = \sqrt{\sum_{i=1}^m \left(\frac{\hat{x}_i - x_{0i}}{s} \right)^2} \quad (2)$$

其中 m 为试验样本容量。均方误的表示与标准均方误类似,只是在式(1)和式(2)中令 $s=1$ 。

由于判别分析仅对恢复对象作分类判别,恢复时,考虑到平均值是最大似然估计量,比中值好些。因此把恢复的类别样本平均值作为类别的恢复值。

分类数和分类方法直接影响类别的平均值。分类数按一般常用的分为3类和5类。考虑到分类后恢复的平均值代表性以类内变化越小越好,因此选取分5类,而且与当前国际对气候要素分类数一致。分类方法一般用正态分类方法^[11],以及目前国际常用的Gamma分布分类法。后者按概率0.1,0.3,0.7和0.9作为5,4,3,2,1类的分界点^[12]。正态分布分类法则把恢复对象按标准化值划分,分5类时,以-1.29,-0.57,0.57,1.29作为5,4,3,2,1类的分界点。

逐步判别与逐步回归都是双重检验选取进入和剔除方程的因子,它们均用 F 统计量进行计算, F 值可以任意选择以控制方程的因子数。考虑到至少有因子可能被选取到, F 值不能太高,本文选取 F 值取1.0~2.5作为试验值。

2 东北地区夏季降水量与气温的恢复方法试验

为了确定选取什么恢复方法和参数,以东北地区 22 个站 50 年资料逐个测站作试验。因为缺测资料常常发生在解放初期,选取 1951 ~ 1955 年的降水量(气温)看作为缺测资料,分别作缺测 1、3 和 5 年的试验,独立检验样本容量 m 分别为 21, 63, 105。取其余 21 个站降水量(气温)作为因子,样本因子矩阵不包含缺测年份,样本容量为 49,用双重检验逐步判别和逐步回归方法选取因子,建立恢复资料的方程。判别时按正态分布和 Gamma 分类。表 1 给出东北地区 22 站夏季降水量不同方法独立试验样本均方误与标准均方误的比较。

表 1 东北地区夏季降水量不同方法试验样本均方误与标准均方误的比较

方法	F 值	1 年		3 年		5 年	
		均方误	标准均方误	均方误	标准均方误	均方误	标准均方误
判别	2.5	140.3	1.30	121.3	1.09	112.9	1.03
判别	2.0	129.4	1.25	114.5	1.07	105.2	0.97
判别	1.5	138.9	1.38	115.7	1.12	113.9	1.06
判别	1.0	151.2	1.46	119.3	1.16	116.5	1.09
回归	2.5	187.5	1.66	162.7	1.43	152.8	1.36
回归	2.0	127.1	1.24	132.3	1.19	139.3	1.26
回归	1.5	131.5	1.21	133.9	1.17	146.4	1.25
回归	1.0	141.5	1.49	143.7	1.31	147.6	1.32

从表 1 中可见,22 站试验均方误比较表明,逐步回归与判别方法恢复 1 年气候要素值,以 F 值取 2.0 有较好的效果。在 3 ~ 5 年恢复试验中,判别比回归效果好。另外,用正态分布划分 5 类,逐步判别 F 值仍然取 2.0,标准均方误在 1 年、3 年和 5 年分别为 1.21、1.10 和 1.03,不如用 Gamma 分布分类好。又考虑到降水量不遵从正态分布,以后的试验,判别方法仅使用 Gamma 分布分类。

对夏季气温进行恢复试验。由于气温分布较好的遵从正态分布,所以不需要作 Gamma 转换。使用判别分析方法时,使用正态分布分类。不同方法试验平均平方根误差比较情况见表 2。

表 2 东北地区夏季气温恢复试验均方误与标准均方误的比较

方法	F 值	1 年		3 年		5 年	
		均方误	标准均方误	均方误	标准均方误	均方误	标准均方误
判别	2.5	0.46	0.60	0.46	0.59	0.52	0.65
判别	2.0	0.41	0.53	0.43	0.56	0.46	0.58
判别	1.5	0.46	0.61	0.45	0.59	0.48	0.62
判别	1.0	0.46	0.61	0.45	0.59	0.48	0.62
回归	2.5	0.45	0.56	0.46	0.59	0.71	0.88
回归	2.0	0.45	0.56	0.46	0.59	0.71	0.87
回归	1.5	0.45	0.56	0.46	0.59	0.70	0.87
回归	1.0	0.45	0.56	0.46	0.59	0.69	0.85

从表 2 中可见,气温的恢复中,回归分析恢复效果比较稳定,随 F 的取值变化不大。可能因为气温较好遵从正态分布,恢复效果较好。但是判别分析方法的恢复在 F 值为 2.0 时有更好的效果。不同方法与夏季降水量的恢复比较中,气温的恢复有更小的误差。

3 东北地区用邻站作夏季降水量与气温的恢复方法试验

考虑到气候要素的地域的均一性,相同地区的气候变化应该是相似的,选择的因子站应该是恢复对象站的邻近站点。因此我们把东北地区的 22 个测站,按地理经纬度转换为直角坐标,用欧氏距离最短的方法选取对象站周围的站点。试验时,考虑到邻站不宜离试验站太远,所以邻站因子站试验数取 3~6 站。由于不是按因子对方程的贡献大小选择因子,而是以邻站选择因子,所以两方法均取 $F=0.0$,即因子全部进入方程,即使用多元回归或多元判别方法进行。不同方法的均方误和标准均方误比较见表 3。

表 3 东北地区 22 站取最近邻站站数恢复试验夏季降水量的平均平方根误差比较

方法	邻站数	1 年		3 年		5 年	
		均方误	标准均方误	均方误	标准均方误	均方误	标准均方误
判别	6	124.9	1.18	108.6	1.00	101.2	0.97
判别	5	146.0	1.30	113.0	1.05	108.5	1.02
判别	4	124.4	1.15	102.2	0.96	107.5	1.00
判别	3	101.9	1.05	96.5	0.90	100.3	0.94
回归	6	138.0	1.33	127.8	1.18	124.5	1.17
回归	5	141.1	1.36	138.6	1.30	139.4	1.33
回归	4	141.0	1.25	132.9	1.18	126.3	1.16
回归	3	124.4	1.17	122.3	1.09	124.6	1.19

从表 3 可见,使用回归方法,选择 3 个邻站的 3 年的降水量恢复中有较小误差,其标准均方误大体比从 22 个站用 F 值选取因子的均方误要小些。使用判别方法,选择 3 个邻站 1~5 年的降水量恢复有较小误差。而且效果比回归分析方法要好,其标准均方误也比从 22 个站用 F 值逐步选取因子的均方误要小。说明邻站使用判别方法恢复方案恢复效果较好。

类似地对夏季气温进行恢复试验,其中使用判别方法时,分类方法直接用正态分布分类。不同方法试验结果见表 4。

表 4 东北地区 22 站取最近邻站站数恢复试验夏季气温的均方误与标准均方误比较

方法	邻站数	1 年		3 年		5 年	
		均方误	标准均方误	均方误	标准均方误	均方误	标准均方误
判别	6	0.44	0.59	0.43	0.56	0.47	0.59
判别	5	0.44	0.59	0.43	0.55	0.46	0.58
判别	4	0.48	0.62	0.46	0.59	0.52	0.64
判别	3	0.50	0.65	0.51	0.65	0.55	0.68
回归	6	0.41	0.52	0.47	0.58	0.91	1.09
回归	5	0.48	0.58	0.55	0.68	0.67	0.83
回归	4	0.79	0.98	0.76	0.95	0.86	1.07
回归	3	0.54	0.70	0.53	0.67	0.63	0.78

从表 4 可见,邻站选择在气温恢复中,回归方法选择 6 个邻站的气温恢复有较小的误差,其标准均方误差大体比 22 个站用 F 值选取因子的要小。与使用判别分析方法比较,虽然在 1 年的恢复中,其效果不如回归方法。但是在选择 5 个邻站的 3~5 年的气温恢复中,比回归方法误差小。而且在 5~6 邻站中,均方误差变化不大,效果也比回归方法稳定。因此确定使用判别方法选取 5 个邻站作为气温恢复的主要方法。

4 用邻站作全国夏季降水量与气温的恢复试验

由上述试验发现,用邻站作夏季降水量与气温的恢复方法试验,有较好的效果。而且邻站的恢复有更好的物理意义。因此用邻站恢复方案,同样地使用上述两种方法对我国其他 5 个气候地区作降水量和气温恢复的试验。根据东北地区的试验结果,还考虑到降水量邻站的均匀性不如气温好,选择 3 个邻站恢复降水量,气温的恢复则选取 5 个邻站进行恢复。由于各区地理位置不同,仅比较它们的标准均方误差。降水量和气温恢复试验结果见表 5 和表 6。

表 5 不同方法对不同地区夏季降水量的恢复试验样本标准均方误差比较

地区	邻站数	1 年		3 年		5 年	
		判别方法	回归方法	判别方法	回归方法	判别方法	回归方法
东北	3	1.05	1.17	0.90	1.09	0.94	1.17
华北	3	1.26	0.84	1.01	0.92	1.10	1.00
黄河流域	3	0.76	0.79	0.81	0.94	0.82	0.91
长江流域	3	0.66	0.82	0.70	0.90	0.80	1.13
华南	3	0.65	0.69	0.66	0.81	0.75	1.11
新疆	3	0.68	0.65	0.93	0.80	0.94	0.89

表 6 不同方法对不同地区夏季气温的恢复试验样本标准均方误差比较

地区	邻站数	1 年		3 年		5 年	
		判别方法	回归方法	判别方法	回归方法	判别方法	回归方法
东北	5	0.59	0.58	0.55	0.68	0.58	0.83
华北	5	0.83	1.64	0.86	1.55	0.90	1.44
黄河流域	5	1.01	1.06	1.02	1.43	0.92	1.50
长江流域	5	0.79	0.95	0.91	1.21	0.78	1.17
华南	5	0.74	1.03	0.85	1.31	0.79	1.33
新疆	5	1.31	1.38	1.10	1.35	1.09	1.51

从表中比较发现,华北地区由于汛期降水量变化大,恢复误差很大,而华南地区降水量变化小,恢复的误差小。气温恢复则由于同样原因,新疆地区气温变化大,恢复误差大。东北地区夏季气温变化小,恢复效果较好。使用判别方法,各地区气温和降水量的恢复值约在 1 倍标准差范围内变化。

用回归方法,作 1~5 年的降水量恢复试验,有 4 个地区恢复效果不如判别方法。作 1~5 年的降水气温恢复试验,除东北地区 1 年试验略好外,其余地区各年恢复效果不如判别方法。因此,使用邻站的多元判别方法有较好的效果,可以在我国夏季降水量和气温

恢复中实际应用。

5 结 论

本文对我国汛期降水量和气温缺测资料及资料均一性恢复的方法进行试验研究。对单个气候地区(东北区),以及全国 6 个气候地区分别用逐步回归和逐步判别方法进行缺测 1~5 年恢复试验。有如下结果:

(1) 对东北地区内所有站点试验表明,使用双重检验 F 值,进行引入和筛选因子,发现逐步判别分析方法比逐步回归分析的方法恢复效果好些。

(2) 对东北地区内所有站选择邻站建立恢复模型,结果发现无论回归分析还是判别分析,恢复效果均比使用 F 值筛选因子的恢复方案好。

(3) 对东北地区内所有站选择邻站建立恢复模型,结果发现判别分析比回归分析恢复效果好。降水量与气温的恢复,邻站数分别选取 3 和 5 有较好的效果。

(4) 对不同地区判别分析恢复试验表明,夏季降水量和气温的恢复,以华南和东北地区有较好的效果。

参 考 文 献

- 1 Peterson T, David R E, Karl T R, et al. Homogeneity adjustments of in site atmospheric climate data: a review. *Int. J. Climatol.* 1998, **18**: 1493 ~ 1517.
- 2 曾红玲,高新全,戴新刚.近 20 年全球冬、夏海平面气压场和 500 hPa 高度场年代际变化特征分析.高原气象, 2002, **21**(1): 66 ~ 73.
- 3 黄荣辉等.我国夏季降水的年代际变化及华北干旱化趋势.高原气象, 1999, **18**(4): 465 ~ 476.
- 4 施能,陈家其,屠其璞.中国近 100 年来 4 个年代际的气候变化特征.气象学报, 1995, **53**(4): 431 ~ 439.
- 5 李崇银,李桂龙,龙振夏.中国气候年代际变化的大气环流形势对比分析.应用气象学报, 1999, **10**(增刊): 1 ~ 8.
- 6 Potter K W. Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Wea. Rev.*, 1981, **109**: 2040 ~ 2045.
- 7 Easterling D R, Peterson T C. A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.* 1995, **15**: 369 ~ 377.
- 8 Young K C. A three-way model for interpolating for monthly precipitation values. *Mon. Wea. Rev.*, 1992, **120**: 2561 ~ 2569.
- 9 黄嘉佑.气象统计分析与预报方法.北京:气象出版社, 2000. 302.
- 10 黄茂怡,黄嘉佑. CCA 对中国夏季降水场的预报试验和诊断结果.应用气象学报, 2000, **11**(增刊): 31 ~ 39.
- 11 中央气象局等.中国近五百年旱涝分布图集.北京:地图出版社, 1980.
- 12 黄嘉佑.一种用于旱涝分析的降水概率指标——Gamma 分布概率指标.气象, 1990, **16**(9): 3 ~ 12.

THE EXPERIMENTAL STUDY OF RECONSTRUCTION FOR SUMMER PRECIPITATION AND TEMPERATURE IN CHINA

Huang Jiayou

(*Department of Atmospheric Science, School of Physics, Peking University, Beijing 100871*)

Liu Xiaoning Li Qingxiang

(*National Meteorological Center, China Meteorological Administration, Beijing 100081*)

Abstract

The examination of reconstruction for missing data and homogeneity for summer precipitation and temperature in China was completed. The experimental study is on the reconstruction for missing data for the period of 1 ~ 5 years using stepwise regression and stepwise discrimination. It is found that the stepwise discrimination is better than the stepwise regression in the validity of the reconstruction. The experimental studies about different methods of discriminate group show that the discrimination based on Gamma distribution is better with a good validity. The discrimination selected nearly stations have a good validity than the stepwise discrimination with F value. The experimental results also show that the reconstruction for missing with the little error in the six areas in China on reconstruction using discrimination selected nearly stations.

Key words: Data reconstruction Homogeneity Stepwise discrimination Stepwise regression