

北京地面气温可预报性及缺测资料恢复的研究*

黄嘉佑

(北京大学地球物理系, 北京, 100871)

摘 要

利用1951—1990年期间北京地面气温资料作未来气温预测可预报性及缺测气温资料恢复的研究,对单月序列使用自回归、选阶自回归、逐步回归和预测残差最小逐步回归及对多月序列使用后两种模型等6种方案进行试验。结果表明,多月序列使用残差最小逐步回归模型有最好的可预报性,且预报方程具有较高的稳定性。本文使用该方案还对1841—1950年期间的缺测气温资料进行恢复。

关键词: 气温,可预报性,资料恢复,残差最小逐步回归。

1 引 言

北京是华北地区具有代表性的测站,其地面气象要素(降水和气温)具有较好的代表性^[1],而且该站有百余年的实测资料,往往被利用作为恢复或延长其他地区资料的基本测站资料,在研究百年尺度气候变化中起十分重要的作用^[2]。北京地区缺测资料也十分严重,在解放前的110a资料中有34a不同月份缺少观测资料。因此较准确估计并恢复缺测的资料具有十分重要的意义。对于缺测资料的恢复,过去常使用邻站未缺测资料进行比较恢复,或者利用相邻年份资料作线性内插来恢复^[2]。但是,在抗日战争时期,相邻测站的资料往往也同时缺测,用此法很难恢复。而利用相邻年份资料作内插法,其估计的准确性是有限的。近年来,有人提出EOF方法^[3]作为恢复缺测资料的方法,这虽然是一较准确的估计方法,但由于缺乏邻站或邻月资料而很难使用。因此,研究利用其自身前后期(年份)演变关系作缺测资料的恢复是比较可行的方案。

这一研究方案与利用前期已有的资料来估计未来的资料,即利用要素自身演变规律作预测是一样的,是属于自身演变的可预报性研究。在可预报性研究方面,地面气候系统的形成及其变化除了考虑外在因素(如大气环流、海温或太阳等)作用外,还应考虑自身的演变。因此,在作地面气象要素的可预报性的研究中,除了研究地面气象要素对外界因素的响应外^[4]还应考察利用气象要素前期信息对后期要素发展变化的影响^[5]。事实上,气象要素前期信息的形成也包含着与之同一时期外界因素相互作用的结果。因此,研究气象要素对前期气象要素的响应,在某种意义上说也是研究前期外界因素及气象要素自身对后期气象要素的影响。因此,这一研究也是气象要素可预报性研究的重要方面。

* 1993年10月4日收到原稿,1994年3月10日收到修改稿。是国家自然科学基金资助项目。

月平均地面气象要素是气候变化和长期预报研究的主要对象之一,对于某个月份的气象要素来说,前期气象要素既可以看成同一月份前期不同年份的气象要素值(单变量前期时刻之值),也可以看成前期不同年份不同月份的气象要素值(多变量前期时刻之值)。前者仅考虑前期年变化的影响,而后者则包含前期年变化和月变化的影响,它们在可预报性上有何差异也是值得研究的。

气温是研究气候变化的重要气象要素之一,本文将用北京 1951—1990 年 12 个月的地面气温资料作该要素的可预报性研究,寻找好的预报方案,然后推广至北京 1841—1950 年期间以便恢复其中缺测资料。

2 单月序列恢复的方案

用单月序列作气温序列的恢复或预测方法一般常用自回归模型(AR)、滑动自回归模型(ARMA)或季节滑动自回归模型(ARIMA)作预测。由于用差分法 ARIMA 可转化为 ARMA,而 ARMA 模型又可转化为自回归模型,我们以 AR 模型为研究的基础;传统 AR 模型是建立序列依次落后的 p 阶预报方程,我们称为 AR1 模型。为改善逐阶的预报关系,选择自相关好的落后的某阶序列,建立跳阶自回归预报方程,例如选择落后阶数有较大的自相关系数,此模型称为 AR2。另外,考虑到 AR 模型要求原序列具有良好的平稳性,这种要求在一般气象要素序列中很难满足。本文提出可用多元回归方程来代替,即预报方程中的因子是由原序列落后不同步长所生成的新变量。显然,新变量的数学期望和方差不再需要假设与原序列相同,而且可用逐步回归进行因子筛选,为了与前面方案比较,逐步回归过程不用 F 检验的值来控制选入的变量,而是用进入方程变量个数来控制,我们把此方案记为 RE1。为进一步提高预报的准确度,还可用预报残差最小的方法进行因子筛选^[6],并把这方案记为 RE2。这一方法是为了寻找在独立样本中有较好的预报效果,用刀切法寻找进入回归方程的好因子,称为预报残差最小的逐步回归方法。它较之传统的逐步回归具有计算简便,适于小型计算机计算等优点。

为了充分利用前期气象要素信息,在作气温某月缺测资料恢复时不仅考虑同月前期年份的影响,还考虑不同月份前期年份的影响,即不仅把同一月份序列转化为前期若干个新变量,而且还在方程中引入不同月份的前期若干个变量,并进行逐步回归,把这方案记为 MRE1。在类似考虑的多元回归方程基础上用预报残差最小的方法进行因子筛选,把此方案记为 MRE2。

上述 6 个方案恢复效果的检验是对逐月资料进行的,由于 1951—1990 年期间有完整的准确观测资料,我们将把其中某一年份各月气温资料假定为缺测资料,用上述方案进行恢复,计算 12 个月的实测值与估计值的均方误差,记为 RMS,作为各方案可预报性大小的度量。

3 单月序列因子的可预报性

仅用单月序列前期作气温序列的恢复试验是取 1951—1980 年各月资料为基本样本,样本容量 N 为 30,以 1981 年各月为预测对象。用上述关于单序列预报的前 4 个方案进行试验,考虑到在计算落后相关系数阶数较大时,作为总体的估计误差较大,所以取最大落

后阶数为 $N/2$ 。试验阶数从 2—15, 在 AR1、RE1 和 RE2 方案中选取因子试验个数也从 2 变化到 15, 得到不同阶数和不同因子数的各个试验方程。在所有方程中选取各月预报最好的独立预报效果进行比较, 即对 1981 年各月气温预报绝对误差值进行比较, 结果列在表 1 中。

表 1 不同方案各月预报最小绝对误差比较

月份	1	2	3	4	5	6	7	8	9	10	11	12	平均
AR1	0.03	0.01	2.40	3.21	1.00	0.82	2.01	0.26	0.66	1.33	1.32	0.82	1.16
AR2	0.04	0.00	0.77	2.96	0.53	0.82	1.74	0.00	0.03	0.30	1.28	1.52	0.83
RE1	0.00	0.01	0.98	0.41	0.05	0.73	0.19	0.09	0.26	0.73	0.00	0.88	0.36
RE2	0.00	0.06	0.05	0.11	0.06	0.80	0.53	0.22	0.11	0.47	0.01	0.81	0.27

从表中可见, RE1 和 RE2 在依赖样本中有较大的解释方差。上述比较结果表明用回归法筛选因子有较好的可预报性。事实上对在 AR2、RE1 和 RE2 方案中选不同因子的所有试验方程的回归解释方差作平均, 所得结果分别为 0.33, 0.51 和 0.44, 所有试验方程中最佳方程对预报年各月作的预报均方误差分别为 1.49, 1.57 和 1.34。可看出回归方法有较好的效果。

4 多月序列因子的可预报性

对于某个月份的气温预报, 前期因子可以取前期不同年份的各月气温来作为前期气候的影响, 这种考虑的预报方案仅能取回归方案, 类似 RE1 和 RE2 的考虑方法, 只不过选取的因子不限于同一月份, 而是前期不同月份的气温变量, 对应方案记为 MRE1 和 MRE2。筛选因子个数的试验以落后步长 1—10a 为试验范围, 30a 为基本样本, 方程中因子个数取 1—10 个作试验。结果发现其效果又比上述方案有所提高。在 MRE1 和 MRE2 所有试验方程中最佳方程对预报年各月作的预报均方误差分别列在表 2 中, 从表中可看见, 此两种方案可预报性较单月序列有明显提高, 所有试验方程的回归解释方差的平均值分别达到 0.85 和 0.56。说明多月回归方案有更好的可预报性, 可作为恢复气温资料的实用方案。

表 2 多月回归方案各月最佳预报方程预报均方误差比较

方案	1	2	3	4	5	6	7	8	9	10	11	12	平均
MRE1	0.03	0.02	0.66	0.02	0.00	0.01	0.00	0.00	0.00	0.01	0.08	0.18	0.08
MRE2	0.04	0.06	0.23	0.25	0.00	0.01	0.02	0.00	0.00	0.01	0.05	0.00	0.06

在用 MRE1 和 MRE2 作恢复气温资料时, 取多少年的前期资料作为建立回归方程的样本呢? 我们取样本容量 N 分别为 20, 25, 30, 35 和 39 的 5 种情况作试验。表 3 给出最佳预报方程预报年各月作的预报均方误差 (RMS) 及所有方程平均解释方差 (MRR) 比较。

表3 不同方案最佳预报方程预报均方误差及平均解释方差比较

样本容量	20		25		30		35		39	
	RMS	MRR	RMS	MRR	RMS	MRR	RMS	MRR	RMS	MRR
MRE1	1.04	0.90	1.37	0.87	1.26	0.85	1.07	0.77	1.06	0.74
MRE2	0.99	0.72	1.17	0.61	1.44	0.59	1.00	0.48	1.07	0.41

从表中可看见,样本容量 N 为 20 有最好的可预报性。表 3 给出两方案各月最佳预报方程预报均方误差比较。比较表 2,可看见此两方案在此样本中有最好的预报效果。事实上,Mooley 等人^[7]在研究印度降水与 700hPa 的关系也发现它们相关系数在 20—25a 的样本容量中有较好的预报稳定性。实际上,对各月气温 40a 资料作谱分析也发现大部分月的主要周期为 20a(见表 4)。用 20a 为样本容量是比较符合气温自身变化规律的。

表4 各月气温主要周期(T1)和次要周期(T2)(a)

月份	1	2	3	4	5	6	7	8	9	10	11	12
T1	4.0	20.0	10.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	3.3	2.2
T2	20.0	2.5	2.2	4.0	4.0	4.0	2.2	4.0	2.2	2.2	10.0	6.7

试验也发现残差筛选回归方案(MRE2)虽然解释方差不如 MRE1 好,但预报的效果略好。因此以此方案为恢复气温缺测资料的主要方案。为了寻找各月合适的稳定恢复回归方程,我们用 1951 年开始的 20a 为试验样本,滑动地作 3a 预报(因为历史缺测资料一般为 2—3a),以 1—10 和 1—10 为落后年(阶)数(L)和因子数(K),取预报 3a 的预报均方误差(RMS)来衡量可预报性是否有稳定效果,表 5 给出各月有 3a 预报均方误差最小的值及对应方程落后阶数和方程中因子数。

表5 各月最小预报均方误差的方程落后阶数和方程中因子数

月份	1	2	3	4	5	6	7	8	9	10	11	12
L	1	8	4	1	1	8	3	10	2	3	9	6
K	5	10	7	10	10	8	10	9	4	3	2	9
RMS	1.30	0.42	0.35	0.98	0.81	0.70	0.55	0.94	0.07	0.15	1.25	0.35

从表中可看见用同一方程连续作 3a 预报除 1 和 11 月外,其余月份均可有小于 1.0°C 的均方误差。这种预测误差无论是对未来的预测还是缺测资料的恢复均是可以接受的,因此,对各月气温缺测资料将用此方案进行恢复。

5 缺测资料的恢复

在 1841—1950 年期间的 110a 北京气温资料中,有完整各月资料的年份有 76a,其余 34a 均需进行恢复,逐年用 MRE2 方案对该年缺测资料进行恢复,为保证恢复的资料有最小的误差,我们用上述方法选取恢复年的前 23a 的资料,分别以 20a 为试验样本,滑动地作 3a 预报(因为历史缺测资料一般为 2—3a),以 1—10 和 1—10 为落后年数(L)和因子

数(K),取预报 3a 的预报均方误差(RMS)来衡量可预报性是否有稳定效果,选取各月有最小预报均方误差的方程落后阶数和方程中因子数为预报方程。对于连续几年缺测资料可使用同一预报方程。如此用 76a 实测资料恢复出 110a 资料。为检验恢复资料的可信性,我们对 76a 和 110a 的 12 个月的两个资料阵 X1 和 X2 作主分量分析。表 6 给出它们前 5 个主分量的解释方差比较。

表 6 两个资料阵前 5 个主分量的解释方差比较

分量	1	2	3	4	5	总和
X1	0.185	0.142	0.108	0.101	0.092	0.628
X2	0.187	0.127	0.112	0.099	0.091	0.616

从表中可看见,它们是十分相似的,特别是第 1 分量,相差仅 0.02,前 5 个分量的累积解释方差仅相差 0.012。这两组主分量的相似系数为 0.999。如果比较它们对应的特征向量的相似性,这 5 个特征向量的相似系数分别为 0.930、0.937、0.886、0.827 和 0.974。可见它们的分布形势也是十分相象的。

由于 110a 的恢复资料阵是可信的,且它的第 1 主分量有最大的解释方差,以此分量为北京气温该期间多年变化的代表。从该分量二项系数滑动曲线在 1881—1950 年时段的变化中发现:1920 年代的气温有一突变;在百年变化尺度中 1940 年代的增温现象表现是十分明显的。这些变化特点与前人工作是一致的。说明本文气温的恢复有一定的可信性。

6 结 论

本文利用 1951—1990 年期间北京地面气温资料作未来气温预测可预报性及缺测气温资料恢复的研究,对单月序列使用自回归、选阶自回归、逐步回归和预测残差最小逐步回归及对多月序列使用后两种模型等六种方案进行试验。结果有如下结论:

(1) 单月气温序列中预测残差最小逐步回归方案有最小的预报误差,误差平均约为 1.3℃,最优时可达 0.27℃。试验还发现春秋转换季节预报误差较其它季节要大些。

(2) 多月气温序列中预测残差最小逐步回归方案仍有最小的预报误差,且较单月序列的要小,误差平均约为 1.0℃,最优时可达 0.06℃。该方案可作为气温缺测资料恢复的主要方案。

(3) 对不同样本容量试验,发现用预测残差最小逐步回归方案,容量为 20a 的样本有最小的预报误差。

(4) 用预测残差最小逐步回归方案对 1841—1950 年期间缺测气温资料进行恢复,恢复后的资料多年变化特点与实测资料及气温级别资料十分相似。

参考文献

- [1] 黄嘉佑. 用概率指标作华北夏季干旱现象分析. 大气科学研究与应用. 1992. (3): 7—15.
- [2] 唐国利等. 1921—1990 年我国气温序列及变化趋势. 气象. 1992. 18(7): 3—6.
- [3] 张邦林等. 用前期大气环流预报中国夏季降水的 EOF 迭代方案. 科学通报. 1991. 23: 1797—1798.

- [4] 黄嘉佑. 长期天气过程的可预报性研究的进展. *气象科技*, 1992, 3: 30—37.
- [5] 丑纪范. 长期数值天气预报. 北京: 气象出版社, 1986.
- [6] 黄嘉佑, 王云漳. 用预报残差最小的逐步回归方法作黄河上游旱涝预测试验. *高原气象*, 1990, 9, 439—442.
- [7] Mooley D A and Paolino D A. Relationship of the Indian monsoon rainfall to the Northern Hemispheric 700 mb height tendency. *J Climatol*, 1988, 8: 499—509.

A STUDY OF PREDICTIVITY AND RECONSTRUCTION FOR MISSING RECORDS FOR TEMPERATURE IN BEIJING

Huang Jiayou

(*Department of Geophysics, Peking University, Beijing, 100871*)

Abstract

Using the data of temperature in Beijing during the period of 1951—1990, the predictivity and reconstruction for missing records is investigated, the prediction equations are built by six statistical models: autoregression, autoregression for selective order, stepwise regression and stepwise regression with minimum of forecast error for uni-month series, and the later two models for multi-month series. The results show that the model of stepwise regression with minimum of forecast error has the best predictivity in all of the models. The reconstruction for missing records during 1841—1950 for temperature in Beijing has been completed using the model of stepwise regression with minimum of forecast error.

Key words: Temperature, Predictivity, Reconstruction stepwise, Regression with minimum of forecast error.