

北京降水缺测资料恢复的研究¹⁾

黄嘉佑 张 鐔

(北京大学地球物理系, 北京 100871)

提 要

利用 1941—1980 年北京降水量资料进行了缺测降水量资料恢复的研究, 对单月序列使用非线性函数作降水量正态化的变换, 然后用预测残差最小逐步回归及对多月序列使用各月降水量变换序列、各月降水量变换序列与各月气温混合序列为因子作缺测资料恢复和用各月缺测资料前后实测资料作简单内插恢复等三种方案进行试验。结果表明, 以降水和水温作混合因子使用残差最小逐步回归模型有最好的可预报性。还使用该方案对 1841—1950 年的缺测降水量资料进行恢复。

关键词: 降水量正态化 非线性变换 残差最小逐步回归

引 言

北京是华北半干旱地带具有代表性的测站, 该站的降水资料对华北地区旱涝气候状态具有较强的代表性^[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, 对研究该地区水资源的长期演变规律, 或在研究百年尺度气候变化中起十分重要的作用^[2]。北京地区缺测资料是十分严重的, 在 1949 年前的 110 年资料中有 34 年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内插, 其估计的准确性是有限的。近年来, 有人提出 EOF 方法^[3]虽然较准确, 但由于缺乏邻站或邻月资料又很难使用。因此, 虽然已有恢复好的北京降水资料, 研究较好的恢复方案仍是十分必要的。本文提出利用其自身前后期(年份)演变关系作缺测资料的恢复。

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)。前者仅考虑前期年际变化的影响, 而后者则包含前期年变化和不同月份的年际变化的影响, 它们在可预报性上有何差异也是值得研究的。

利用前期气候状态作降水量的恢复, 自然希望它们的关系比较稳定, 即处在非突变的气候阶段。不少研究表明, 地球气候变化似乎在 1980 年前后有一百年尺度的转折, Chiu 发现南半球冰盖长期变化在 1980 年有一转折^[4], 张鐔等人发现全球海平面气压场第一主分量的百年演变中 1980 年附近也有一转折^[5], 黄嘉佑对西太平洋副热带高压的长期演变规律研究中也发现 1980 年是一转折年份^[6], 可见至 1980 年为止的近百多年期间为一气候较为稳定阶段, 又考虑到短期气候阶段多取 30—40 年为一阶段。本文用北京 1941—1980 年 12 个月的降水量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 年以便恢复其中缺测资料。

1 预报恢复缺测值的模型

根据上述理由, 某月份降水量距平的变

1) 本研究得到国家基础研究重大关键项目和自然科学基金资助

化可表示为由前期同一月份和不同月份的降水量和气温线性关系所影响,例如某 t 年1月降水量可写作

$$\begin{aligned} R_1(t) = & a_1 R_1(t-1) + a_2 R_1(t-2) \\ & + \cdots + a_p R_1(t-p) + b_1 R_2(t-1) \\ & + b_2 R_2(t-2) + \cdots + b_p R_2(t-p) \\ & + \cdots + c_1 R_{12}(t-1) + c_2 R_{12}(t-2) \\ & + \cdots + c_p R_{12}(t-p) + d_1 T_1(t-1) \\ & + d_2 T_1(t-2) + \cdots + d_p T_1(t-p) \\ & + f_1 T_2(t-1) + f_2 T_2(t-2) + \cdots \\ & + f_p T_2(t-p) + \cdots + g_1 T_{12}(t-1) \\ & + g_2 T_{12}(t-2) + \cdots \\ & + g_p T_{12}(t-p) + e(t) \end{aligned} \quad (1)$$

式中 R_i 和 T_i ($i=1,2,\cdots,12$)分别表示降水量和气温, p 为所考虑的前期年份数,它们代表前期大气演变的状态, a,b,c,d,f 和 g 分别为对应组合项的系数, $e(t)$ 代表不能被前期气象要素所解释的部分。这一预测某年某月降水量距平模型既包括传统的自回归模型,也包括不同因子的回归模型,还应考虑系数的时变性,这是一个复杂的自适应线性模型。如果假定在短期内大气相互影响关系不会发生太大的变化,在利用上述模型恢复某一年份的资料时可把系数看作不变,在时段改变时再重新建立预报模型。因此,在短期内上述模型可使用多元回归建模的方法定出模型中的系数。

在确定预报模型中,由于因子数很多(如选前5年气候状态作因子,则有120个待选因子),必须进行筛选,即一般用拟合残差最小的最小二乘法基础上的逐步回归算法定出。本文使用预报残差最小的逐步回归方法以每步只引入一个因子为主要计算方案。把预报值与实况值之差称为预报残差,以使它为最小的原则进行因子筛选和建立预报模型^[7]。

2 单月降水序列的正态性转换

用模型(1)对缺测月降水量恢复和预测,但是试验表明,恢复效果很差。究其原因,主

要是由于降水量变差很大且又不遵从正态,用回归方程预测不易取得理想效果。我们采取几种方案来克服这些缺点,分别使用逐月序列标准化和0—1化的方案使降水量的变化减小,计算表明转换后序列有接近于1的均方差,可以克服降水量变差大的缺点。但是,因为这些转换是线性变换,并不改变分布结构,未能改善非正态性的缺点。因此我们考虑对序列作非线性变换。

为寻找序列合适的非线性变换,我们对常用的指数、自然指数和对数等三种类型的非线性函数作试验,设降水量为 x ,变换后的变量为 y ,函数中的常数 a 可用最优化方法确定,最优化的目标函数是使变换后的变量具有接近零的偏度系数,即使其较好的遵从正态分布。试验表明指数函数的变换可以有最小的偏度。表1列出转换前序列逐月转换后序列 x 和 y 的偏度和峰度系数的比较,从表中可见,逐月转换前序列 x 的偏度和峰度系数均明显大于零。表明该变量是非正态的。而转换后序列 y 的偏度和峰度系数均接近于零,尤其是偏度系数。因此说明该变换能满足上述要求。表中还列出转换函数的指数值,它的变化均小于0.5,1月和12月份还出现负值。这就表明,在我国北方地区,一般所认为的用平方根函数对降水量作转换并不是最好的。

表1 转换前后序列 x 和 y 的偏度和峰度系数的比较

月份	x 偏度系数	x 峰度系数	y 偏度系数	y 峰度系数	a 指数
1	3.696	17.753	0.000	-1.297	-0.1016213
2	1.916	3.910	0.000	-1.082	0.2279317
3	2.584	9.904	0.000	-0.591	0.2768263
4	3.121	13.305	0.000	-0.150	0.2866043
5	1.581	2.658	0.000	-0.128	0.2852599
6	1.636	3.086	0.000	0.060	0.2707729
7	1.849	5.415	0.000	0.858	0.2999024
8	1.241	1.710	0.000	-0.257	0.2326316
9	1.798	4.789	0.000	0.244	0.3878837
10	2.236	7.626	0.000	-0.663	0.3043356
11	4.611	28.765	0.000	-0.170	0.2631558
12	2.755	8.954	0.000	-1.496	-0.0544607

3 降水量序列的恢复试验

仅用单月序列前期作降水量序列的恢复试验是取 1941—1980 年各月资料为基本样本,样本容量 N 为 40,以 1981 年各月为预测对象。

对于某个月份的降水量预报,前期因子可以取前期不同年份的各月降水量来作为前期气候的影响,我们称为方案 1,另外,前期因子还可以取前期不同年份的各月降水量和气温来作为前期气候的影响,我们称为方案 2。

在用残差筛选回归方案恢复降水量资料时,我们取样本容量 N 分别为 20、25、30、35 和 39 的 5 种情况对降水量指数转换作试验,例如取样本容量 N 为 20 时,分别用 1941—1960 资料作 1961 年各月降水量的转换值预报,用 1941—1965 资料作 1966 年各月降水量的转换值预报等。预报时,为减少因子转换值的不同方差影响,进一步对转换值作标准化处理。考虑到在计算落后相关系数当阶数较大时,总体的估计误差较大,所以取试验阶数从 1—5,选取因子试验个数从 1 变化到 10,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,即对预报年各月降水

量预报均方误差进行比较,从中选出最佳预报方程。试验表明,方案 1 在 $N=20$ 和方案 2 在 $N=39$ 时有较小的预报平均均方误差,但方案 2 在 $N=20$ 时也有较小的预报平均均方误差。考虑到缺测资料间隔期一般不足 40 年,而且北京地面气象要素多存在 20 年变化的周期性。实际上,对各月降水量 40 年资料作谱分析也发现大部分月的主要周期为 20 年。

因此,用 20 年为样本容量是比较符合降水量自身变化规律的。事实上,Mooley 等人^[8]在研究印度降水与 700hPa 的关系时也发现相关系数在 20—25 年的样本容量中有较好的预报稳定性。使用方案 2,选取 $N=20$ 作为恢复的样本容量是合适的。

由于缺测资料在 1855 年就出现,恢复试验一开始取 $N=14$ 的样本,找寻最佳的预报方程作 1855 年各月预报,以后遇到缺测资料则用前期 20 年作基本样本。以恢复 1884 年的各月缺测资料为例,选取 1863—1882 年逐月的降水量和气温为待选因子,其数目一共是 120 个。以 1983 年各月降水量作预报效果检验。表 2 给出各月预报最佳方程的落后阶数、入选因子个数、与实况(标准化值)的差值比较。

表 2 各月预报最佳方程

月 份	1	2	3	4	5	6	7	8	9	10	11	12
因子数	3	4	7	5	2	9	8	5	9	6	9	8
阶 数	2	2	4	1	5	5	1	2	1	5	4	1
误 差	-0.015	-0.040	0.011	0.144	-0.085	-0.004	-0.086	0.014	-0.001	-0.131	-0.174	0.019

从表中可见,由于标准化值基本变化在 $-3.0-3.0$ 之间,而它们年均方误差仅为 0.0843,说明用此法恢复资料有一定效果。从各月误差值比较而言,4、10 和 11 月份误差较大,超过 0.1。这可能是春季和秋季天气变

化较剧烈的缘故。不过,北京地区湿期为 5—9 月份,该期间效果较满意。

为比较起见,我们还作内插式的恢复试验,即对各月缺测资料用邻近有观测的年份内插出预报值。称此为方案 3。

表3 1841—1950年北京逐月恢复的降水量

年份	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	总量
1841	14	74	265	63	291	2051	2061	1376	743	153	92	9	7192
1842	90	49	66	13	470	730	2533	1862	925	315	0	56	7109
1843	0	283	0	39	869	955	1467	2245	648	34	51	0	6591
1844	15	0	160	229	710	662	2960	1112	235	158	29	3	6273
1845	66	36	8	9	154	927	4440	2010	181	275	4	0	8110
1846	5	45	0	5	267	1245	3258	501	615	91	13	0	6045
1847	0	0	89	38	346	1230	1320	1706	80	103	5	52	4969
1848	23	19	3	562	63	625	5040	1232	214	171	1	11	7964
1849	5	80	35	508	502	1231	1556	1190	1732	456	0	11	7306
1850	5	130	36	147	1072	1524	938	1651	54	31	196	198	5982
1851	31	0	193	173	462	247	2853	877	684	142	15	26	5703
1852	0	181	81	44	333	465	2703	1445	417	102	117	181	6069
1853	51	13	15	285	96	1443	1931	2733	2916	49	317	23	9872
1854	295	81	18	29	206	140	607	1079	879	3	5	0	3342
1855	0	0	23	14	163	460	1951	2553	698	0	241	0	6103
1856	6	0	22	226	122	662	1713	2267	581	111	72	4	5786
1857	15	17	64	73	302	436	2232	2502	308	93	5	2	6049
1858	149	0	72	41	317	329	2057	1836	755	43	239	0	5838
1859	3	2	64	15	225	482	1924	1832	1353	58	224	0	6182
1860	0	32	253	210	1103	666	1085	424	359	353	138	0	4623
1861	42	43	49	1	918	434	2356	1111	244	68	180	9	5455
1862	0	12	0	0	227	412	2050	2255	1054	47	112	15	6129
1863	0	1	69	206	321	516	1369	2336	1045	34	24	2	5923
1864	0	7	113	78	327	528	1331	1807	252	50	14	3	4510
1865	45	6	128	131	317	347	2272	1647	644	87	86	0	5710
1866	1	1	31	63	291	465	964	1332	357	64	4	1	3574
1867	3	0	5	149	294	1031	351	1392	1636	39	9	3	4912
1868	44	2	107	9	353	301	1492	1568	536	30	22	0	4464
1869	0	0	30	62	73	1099	68	252	434	171	227	4	2420
1870	0	0	138	128	172	1863	2061	910	284	204	51	35	5846
1871	1	0	0	200	65	794	3634	3022	2437	397	45	49	10644
1872	10	88	165	173	335	887	3416	713	1006	66	52	0	6911
1873	0	21	0	273	512	96	3638	3725	123	330	0	4	8722
1874	0	0	0	19	493	789	2688	1245	883	360	218	1	6696
1875	0	88	115	5	30	322	2329	514	785	121	4	3	4316
1876	0	16	2	0	57	34	2702	2943	397	32	29	19	6231
1877	79	56	179	438	563	318	475	2395	271	36	100	0	4910
1878	22	3	0	427	512	939	1444	3663	664	402	36	25	8137
1879	17	3	0	176	71	814	3497	1742	1014	218	56	0	7608
1880	82	48	49	151	158	1295	1299	338	346	2	0	0	3768
1881	1	32	2	571	133	1352	735	2281	740	50	121	5	6023
1882	11	48	14	93	337	692	1688	1679	1170	275	183	5	6195
1883	31	148	112	40	464	56	6314	1880	418	285	91	0	9839
1884	10	0	136	26	163	13	2324	1879	745	290	79	0	5665
1885	0	0	50	37	89	113	1459	1551	235	331	115	0	3980
1886	16	3	11	242	31	102	533	1089	277	164	4	0	2472
1887	0	73	0	32	105	194	1501	2013	543	32	13	0	4506
1888	0	40	39	170	190	190	441	1898	315	186	302	0	3771
1889	2	13	90	216	25	532	2714	2428	1194	305	160	0	7679
1890	0	15	5	93	210	145	8718	895	15	266	70	0	10432
1891	0	20	0	118	0	3205	9957	683	0	25	3	0	14011
1892	138	8	33	193	196	985	4473	1894	728	14	189	0	8851
1893	79	0	66	235	612	2373	5827	1392	896	65	83	0	11628
1894	74	4	103	353	148	2817	3226	2441	540	229	137	19	10091
1895	12	68	77	555	49	107	1340	661	170	621	0	37	3697

续表

年份	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	总量
1896	7	3	183	332	318	860	2945	1630	457	65	12	27	6839
1897	160	0	272	578	119	589	3722	703	420	136	44	0	6743
1898	10	87	49	81	10	559	2570	1290	721	5	189	0	5571
1899	42	159	296	217	231	868	1287	231	0	0	132	47	3510
1900	5	54	18	629	131	1749	1023	845	104	11	184	0	4753
1901	10	0	18	176	50	869	2579	912	117	10	200	22	4963
1902	0	16	55	288	55	633	4760	1338	743	15	138	35	8076
1903	6	13	135	331	40	374	889	1661	1099	124	57	0	4729
1904	2	105	75	417	124	1454	4371	2093	641	179	93	3	9557
1905	54	9	345	127	483	969	1109	920	360	443	0	1	4820
1906	40	10	250	7	470	462	2456	1827	575	544	1	0	6642
1907	2	8	1	350	32	130	2775	799	356	417	90	12	4972
1908	491	81	5	21	42	579	2305	1363	1717	156	11	2	6773
1909	6	25	35	74	123	912	5854	464	28	78	30	0	7629
1910	46	0	50	159	129	1176	2606	1392	393	93	233	0	6277
1911	208	0	689	241	109	395	2167	3015	440	45	206	9	7524
1912	145	7	326	173	156	443	1539	1260	25	273	81	0	4428
1913	2	5	223	107	91	537	4213	1116	87	306	22	17	6726
1914	14	104	398	0	90	1169	2562	864	333	958	710	8	7210
1915	29	44	25	5	580	896	4510	824	332	192	52	42	7531
1916	16	60	27	151	283	464	813	1787	327	84	210	74	4296
1917	2	2	26	0	199	377	4383	1201	1224	385	0	18	7817
1918	9	8	149	210	1165	677	1129	1420	50	6	269	14	5106
1919	37	0	80	36	239	911	2641	878	264	521	40	16	5663
1920	28	55	54	10	126	424	1436	271	246	9	39	69	2767
1921	6	0	40	4	352	449	989	621	89	11	1	0	2562
1922	10	40	33	112	82	1351	2979	3374	358	5	32	3	8379
1923	3	3	164	119	35	86	1378	1562	356	37	11	41	3795
1924	4	60	29	0	197	547	6483	2536	375	213	151	71	10666
1925	0	0	27	432	610	1338	4482	1972	640	9	60	0	9570
1926	0	22	22	135	381	716	1109	551	284	4	98	16	3338
1927	198	5	113	71	151	360	3296	1330	145	3	0	46	5718
1928	157	0	74	221	195	640	1112	1818	780	120	120	83	5320
1929	174	0	0	30	100	230	3463	2960	121	153	0	288	7519
1930	0	268	0	69	560	513	1061	1132	288	599	1	24	4515
1931	23	22	56	153	514	435	1536	1994	484	5	22	117	5361
1932	6	9	0	74	406	770	3999	421	1175	2	12	0	6874
1933	4	6	157	324	602	2821	1936	1160	508	41	118	52	7729
1934	12	128	28	100	463	1616	1164	2018	992	55	1	32	6609
1935	10	0	30	0	200	655	1252	1394	15	110	185	0	3851
1936	15	210	0	86	301	375	1156	534	1268	0	69	55	4069
1937	6	43	168	221	180	523	2010	2734	854	66	55	0	6860
1938	0	0	93	80	1393	588	3200	1153	210	626	0	136	7479
1939	0	59	82	89	187	535	3754	1512	136	235	210	7	6806
1940	10	3	1	130	244	741	1771	1177	457	2	1132	48	5716
1941	31	56	14	8	69	1149	564	889	569	55	89	53	3546
1942	1	27	17	256	94	303	2479	1230	173	113	83	0	4776
1943	38	1	11	162	622	107	479	1815	1266	15	405	60	4981
1944	2	14	41	273	307	593	2308	699	423	12	88	0	4760
1945	4	29	64	221	342	247	628	2598	820	10	146	20	5129
1946	28	0	383	406	137	242	3673	1174	611	344	186	6	7190
1947	98	122	93	0	123	409	1949	1694	1412	8	0	116	6024
1948	2	65	44	334	852	490	2187	1007	172	54	103	30	5340
1949	0	55	33	65	576	1127	4173	2078	687	168	72	176	9210
1950	0	65	6	1324	532	585	2891	3344	201	69	86	6	9109

4 缺测资料的恢复效果的比较

由于缺测资料不可能知道,原则上很难比较3种方案的恢复效果。我们只好以总体均匀性进行比较,把去掉缺测资料的资料阵为比较目标。对3种方案所恢复得到的12个月140年(1841—1980)的资料阵(分别记为 R_1 、 R_2 和 R_3)与去掉缺测的资料阵 R_0 作主分量分析,比较它们前几个主分量的特征及其代表性。从中发现,方案3有较大的解释方差,这可能是因为内插资料有较好的均匀性,它平滑了资料的个别变化。与之相反,方案2则有较小的解释方差,它的方差递减规律十分接近于 R_0 阵的递减规律,而且在第一特征场上也比其它的更相似。因此方案2似乎恢复效果较好。说明本文降水量的恢复有一定的可信性。表3给出1841—1950年逐月恢复的降水量。

5 结论

本文利用1951—1980年北京降水量资料作未来降水量预测可预报性及缺测降水量资料恢复的研究,对单月序列使用非线性函数作降水量的变换能减少预测误差。本文还对不同的前期因子,即使用降水量自身为因

子和用降水量与气温混合为因子,分别用预测残差最小逐步回归方法的方案进行恢复比较试验。结果表明与气温相结合的方案有较小的预测误差和较好的恢复效果。

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, (三), 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 Chiu, L. S., Variation of Antarctic sea ice: An update. Mon. Wea. Rev., 1983, 111: 578—580.
- 5 Tan Chang and T. Yasunari. Fluctuations of global surface pressure patterns during the past 100 years and their relation to the Asian Monsoon. Part I. northern summer (July), J. Meteor. Soc. Japan, 60, 1132—1142.
- 6 Huang Jiayou and Wang Shaowu. Investigations on variation of the Subtropical High in the Western Pacific during historic times. Climatic Chang, 1985, 7: 427—440.
- 7 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.
- 8 Mooley, D. A. and D. A. Paolino. Relationship of the Indian monsoon rainfall to the Northern Hemispheric 700 mb height tendency. J. Climatol., 1988, 8: 499—509.

A Study of Missing Data Reconstruction and Predictability for Precipitation in Beijing

Huang Jiayou Zhang Tan

(Department of Geophysics, Peking University 100871)

Abstract

A study of missing data reconstruction and predictability for precipitation in Beijing is completed in the period of 1841—1980. The new variable of precipitation with nonlinear transform has been done for following Gauss distribution. Further the method of stepwise regression with minimum of forecast error has been employed for reconstructing the missing data using the data of temperature and precipitation in earlier months and years as predictors. Three experimental schemes of using only precipitations or the composite of temperature and precipitation and sample insert value between data are adopted. The results show that the scheme of using the composite factors of temperature and precipitation has good predictability. The missing data in the period has been reconstructed.

Key Words: precipitation Gauss distribution nonlinear transform stepwise regression with minimum of forecast error