第 22 卷第 3 期　　南 京 气 象 学 院 学 报　　Vol. 22 No. 3
1999 年 9 月　　Journal of Nanjing Institute of Meteorology　　Sep. 1999

气象场序列几种插补方案的对比试验

江志红,　　裕国,　　屠其璞
(南京气象学院环境科学系, 南京　210044)

摘要: 采用基于 EOFS 的主分量回归(PCR)、EOFS 迭代法(I-EOF)和基于主分量典型相关的典型变量回归(CVR)3 种不同的统计插补计算方案, 对同一区域同一种气象要素序列进行缺测资料的插补试验。结果表明, 各种方案插补精度都与参数选择有关, 无论缺测站点空间分布类型如何, 当缺测点数小于 60 %时, 3 种方案均有较好效果, 以 CVR 最佳, 且随缺测年数增长, CVR 优势更显著。

关　键　词: 气象场序列; 资料序列插补; 典型相关分析; 经验正交函数
中图分类号: P468.01　　文献标识码: A

研究区域性气候的长期变化规律, 须具有连续均一且有相当精度的气象要素场序列。我国较为完整的观测站网基本上是 50 年代初形成的, 此前的观测记录几乎都因种种缘故而有不同程度的缺损, 而缺测站点的空间分布又极不均匀。为了获得较为完整的连续而均一且有相当精度的气象要素观测序列, 利用某些多元统计技术, 在已有的稀疏网点长期观测序列基础上, 较为精确地插补并延长气象场记录序列具有十分重要的意义。

采用统计方法插补延长气象观测记录早在 50 年代即已流行[1,2]。么枕生(1963)最早在国内提出气象观测记录序列订正的问题, 并系统地阐述了这一论题的观点[3]。随着气候变化研究的深入, 借助于更为复杂的多元统计方法, 单一测站资料插补订正思想推广到了整个气象要素场资料序列的插补延长, 在 70～80 年代得到了长足进展。许多学者对此作过有益的探讨, 并取得良好的效果[4～6]。本文采用 EOF 主分量回归(PCR)、EOF 迭代(I-EOF)和典型相关意义下的典型变量回归(CVR) 3 种插补方法, 对长江流域 50 个测站 1951～1990 年 1 月地面气温场序列, 作各种假设缺测情况下的插补延长试验, 寻求最优方案, 以便为建立中国各区域近百年连续、均一的月平均气温序列提供高精度的统计插补方案。

1　插补方案简介

假设要素场 $\boldsymbol{X}=(x_1, x_2, \cdots, x_M)$ 由两个子场 $\boldsymbol{X}_1=(x_1^{(1)}, x_2^{(1)}, \cdots, x_{M1}^{(1)})$ 和 $\boldsymbol{X}_2=(x_1^{(2)}, x_2^{(2)}, \cdots, x_{M2}^{(2)})$ 构成。其中 M_1、M_2 分别为 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$ 的测站数, $M=M_1+M_2$。若已知 $\boldsymbol{X}_1$ 在 $t=1, 2, \cdots, n, n+1, \cdots, N$ 有观测序列, $\boldsymbol{X}_2$ 仅在 $t=n+1, n+2, \cdots, N$ 有观测序列。为叙述方便, 可假设 $\boldsymbol{X}$ 及其子场 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 都已零均值化。

收稿日期: 1999-01-18; 修订日期: 1999-03-28
基金项目: 九五项目 "96-908-01-01" 课题资助
作者简介: 江志红(1963-), 女, 博士生, 南京气象学院副教授

1.1　PCR 方案[4]

由 EOFs 展开, 可分别获得时段 $t = n+1, n+2, \cdots, N$ 的观测序列主分量

$$\boldsymbol{U} = \boldsymbol{E}^{\mathrm{T}}\boldsymbol{X}, \qquad \boldsymbol{U}_1 = \boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{X}_1。 \tag{1}$$

其中 $\boldsymbol{E} = (e_1, e_2, \cdots, e_p)$ 和 $\boldsymbol{E}_1 = (e_1^{(1)}, e_2^{(1)}, \cdots, e_p^{(1)})$ 为 $\boldsymbol{X}$ 和 $\boldsymbol{X}_1$ 的前 p 个特征向量; $\boldsymbol{U} = (U_1, U_2, \cdots, U_p)$ 和 $\boldsymbol{U}_1 = (U_1^{(1)}, U_2^{(1)}, \cdots, U_p^{(1)})$ 为其相应主分量。则可建立主分量回归估计方程

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{U}_1。 \tag{2}$$

在特征向量场稳定的一般条件下, 由子场 $\boldsymbol{X}_1$ 在 $t = 1, 2, \cdots, n$ 的观测序列, 可得相应主分量估计式

$$\hat{\boldsymbol{U}}_1(t) = \boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{X}_1。 \tag{3}$$

由此可得要素场 $\boldsymbol{X}$ 在 $t = 1, 2, \cdots, n$ 上的估计序列

$$\hat{\boldsymbol{X}} = \boldsymbol{E}\boldsymbol{U} = \boldsymbol{E}\boldsymbol{B}\boldsymbol{U}_1 = \boldsymbol{E}\boldsymbol{B}\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{X}_1。 \tag{4}$$

值得指出的是, (4) 式成立的前提是 $\boldsymbol{E}$ 和 $\boldsymbol{E}_1$ 不随观测年限长短而变[9]。显然, 插补精度与参数 P (特征向量场个数) 有关。

1.2　EOF 迭代方案(I-EOF)

将上述 $\boldsymbol{X}$、$\boldsymbol{X}_1$ 写为相应分块矩阵, 就有[7]

$$\boldsymbol{X}^{(0)} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{0} & \boldsymbol{X}_{22} \end{pmatrix}。 \tag{5}$$

其中 $\boldsymbol{X}_{11}$和 $\boldsymbol{0}$ 分别为 $n \times M_1$ 矩阵和 $n \times M_2$ 矩阵, 后者为待插补矩阵; $\boldsymbol{X}_{12}$和 $\boldsymbol{X}_{22}$分别为 $(N-n) \times M_1$ 和 $(N-n) \times M_2$ 矩阵。由于缺测资料序列等价于 $\boldsymbol{X}$ 阵中的分块阵 $\boldsymbol{0}(n \times M_2)$, 假如采用对 $\boldsymbol{X}^{(0)}$赋初值后的逐步 EOFs 展开, 则可利用迭代方法内插分块阵 $\boldsymbol{0}$, 达到插补延长资料的目标。对(5) 式作 EOFs 展开, 并取截断阶数 $K^{(0)}$, 则有 $\boldsymbol{X}^{(0)}$ 的拟合场

$$\hat{\boldsymbol{X}}^{(0)} = \boldsymbol{E}^{(0)}\boldsymbol{U}^{(0)} = \begin{pmatrix} \hat{\boldsymbol{X}}_{11}^{(0)} & \hat{\boldsymbol{X}}_{12}^{(0)} \\ \hat{\boldsymbol{X}}_{21}^{(0)} & \hat{\boldsymbol{X}}_{22}^{(0)} \end{pmatrix}。 \tag{6}$$

其中 $\boldsymbol{E}^{(0)}$和 $\boldsymbol{U}^{(0)}$ 分别为第零步迭代的特征向量阵和主分量矩阵。继续构造矩阵

$$\boldsymbol{X}^{(1)} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21}^{(0)} & \boldsymbol{X}_{22} \end{pmatrix}, \tag{7}$$

对 $\boldsymbol{X}^{(1)}$再行 EOFs 展开。得拟合场

$$\hat{\boldsymbol{X}}^{(1)} = \boldsymbol{E}^{(1)}\boldsymbol{U}^{(1)} = \begin{pmatrix} \hat{\boldsymbol{X}}_{11}^{(1)} & \hat{\boldsymbol{X}}_{12}^{(1)} \\ \hat{\boldsymbol{X}}_{21}^{(1)} & \hat{\boldsymbol{X}}_{22}^{(1)} \end{pmatrix}。 \tag{8}$$

在新拟合场基础上, 继续仿(7) 式构造下一轮矩阵, 并可得相应 EOFs 拟合场。循此往复, 直至第 S 步得到

$$\boldsymbol{X}^{(S)} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21}^{(S-1)} & \boldsymbol{X}_{22} \end{pmatrix}, \tag{9}$$

及相应 EOFs 拟合场

$$\hat{\boldsymbol{X}}^{(S)} = \begin{pmatrix} \hat{\boldsymbol{X}}_{11}^{(S)} & \hat{\boldsymbol{X}}_{12}^{(S)} \\ \hat{\boldsymbol{X}}_{21}^{(S)} & \hat{\boldsymbol{X}}_{22}^{(S)} \end{pmatrix}。 \tag{10}$$

若记 $\hat{\boldsymbol{X}}_{21}^{(S-1)} = (\boldsymbol{X}_{ij}^{(S-1)})_{n \times M_2}$, $\hat{\boldsymbol{X}}_{21}^{(S)} = (\boldsymbol{X}_{ij}^{(S)})_{n \times M_2}$, 则当

$$\max \left| \boldsymbol{X}_{ij}^{(S)} - \boldsymbol{X}_{ij}^{(S-1)} \right| < \epsilon。 \tag{11}$$

时, $\hat{\boldsymbol{X}}_{21}^{(S)}$ 即为前 S 步迭代的精确结果, 其精度参数为 ϵ。显然, $\hat{\boldsymbol{X}}_{21}^{(S)}$中的各元素就是前 n 年子场

$\boldsymbol{X}_2$ 的插补延长序列, 其测站数为 M_2 个。

1.3 CVR 方案

根据 Barnett 等[8] 提出的主分量典型相关分析思想, 利用 $\boldsymbol{X}$ 和 $\boldsymbol{X}_1$ 在时段 $t = n+1, n+2, \cdots, N$ 的典型相关分析结果, 建立典型相关变量的回归模式[9], 即可得到

$$\hat{\boldsymbol{X}} = \boldsymbol{H}\Lambda\Gamma\boldsymbol{E}^{-1}\boldsymbol{X}_1。 \tag{12}$$

其中 $\boldsymbol{H} = \boldsymbol{X}\boldsymbol{V}(t)$ 是 $\boldsymbol{X}$ 场与其典型相关变量 $\boldsymbol{V}(t)$ 的协方差矩阵, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{q^*})$ 为典型相关系数, Γ 是典型相关变量 $\boldsymbol{V}(t)$ 的权重系数向量。正如文献[9]所述, 模式(12)的插补精度还取决于由 $\boldsymbol{X}$、$\boldsymbol{X}_1$ 提出的主分量个数 q、p 及其典型相关变量数 q^*。

2 插补试验区域选取及其精度指标

选用均匀分布在我国长江流域的 50 个测站(图 1)的 1 月平均气温场作为插补试验场, 资料年代为 1951～1990 年。为了比较不同缺测情况下, 各种方案的插补效果, 分别取缺测站点呈插花型分布(SD)和成片型分布(PD), 进行不同缺测站点数(例如 10、20、…)情况下的插补试验, 以确定最优插补方法。图中"1"表示第 1 次试验缺测站点的位置, 第 2 次试验缺测站点位置以"1"、"2"表示, 以此类推。

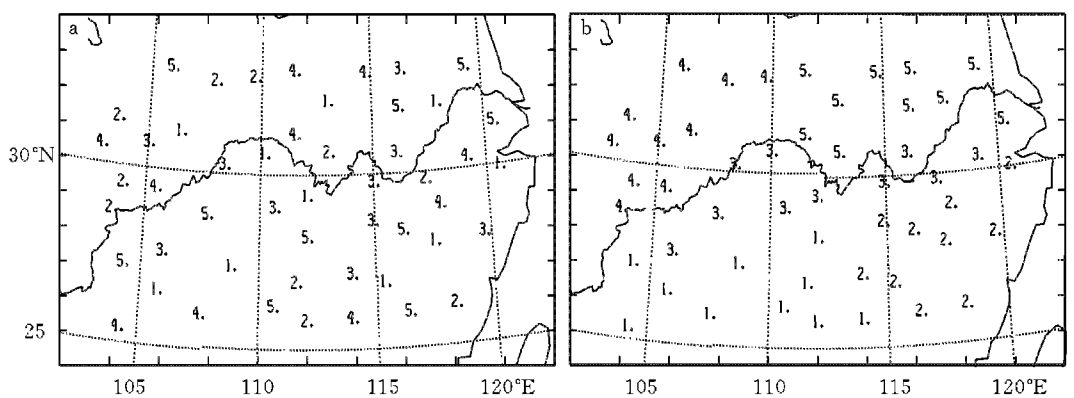

图 1 试验区域及其缺测站点位置分布

a. SD 型; b. PD 型

Fig. 1 The experiment domain and station locations of missing data

a. scatter distribution; b. partial distribution

用前文介绍的 3 种方法, 对上述缺测场进行插补试验, 分别以插补场的均方误差 s、插补距平场与实测距平场的相关系数 r 等统计量, 作为检验插补效果的精度指标。

3 各种方案的参数选择试验

由于 3 种插补方法的精度都与参数的选择有关, 参照文献[9], 取同时段实测样本长度为 25 年, 滑动地对 1960～1951 年 1 月平均气温场作假设缺测 1 年的插补试验, 缺测场取 SD 型分布, 缺测站点数 $M_2 = 20$, 考察不同参数下各种方案的插补精度变化, 以确定最优参数。

3.1 PCR 方案参数的选择

表 1 给出了由 PCR 方法得到的 1960～1951 年 1 月平均气温场插补效果随参数 P(特征

向量数) 的变化状况。可见 $P=5$ 时, 插补效果最好。当然, 对不同缺测场, 参数 P 的选择可略有不同, 但其对精度的影响不是很大。

表 1　PCR 方案平均插补效果随参数 P 的变化

Table 1　Mean interpolation precision as a function of parameter p in PCR scheme

	P					
	2	3	4	5	6	7
s/℃	0.563	0.541	0.535	0.534	0.547	0.546
r	0.802	0.815	0.828	0.833	0.827	0.824

3.2　I-EOF 方案参数的选择

用 I-EOF 方案对相同的缺测场进行大量的插补试验表明, 当迭代精度 $\epsilon=0.05$ 时, 其插补效果已相当稳定(图略)。例如, 当 ϵ 取 0.2、0.1、0.05、0.02、0.01 时, 1960～1951 年 1 月气温插补场的平均精度(以距平场相关系数为指标) 分别是 0.790、0.810、0.816、0.816、0.816, 故本文以下有关试验统一取迭代精度 $\epsilon=0.05$。显然, 每次迭代时截取的特征向量数 $K^{(i)}$ 也会影响插补精度。为方便起见, 各次迭代取统一截断数 K。图 2 给出了 1960～1951 年平均插补效果随截断阶数 K 的变化, 显然, $K=3$ 时, 精度最高。对不同缺测场的试验表明, 平均插补效果随参数 K 的变化不大。根据文献[9], CVR 方案的最优参数分别为 $p=10, q=11, q^{*}=9$。

4　插补精度对比分析

4.1　不同缺测点分布类型的插补效果

在最优参数配合下, 用上述 3 种方法, 分别作 1 年缺测的插补滑动试验。缺测年份依此为 1960～1951 年。表 2 和表 3 分别列出缺测场呈 SD 和 PD 分布的插补结果精度对比。

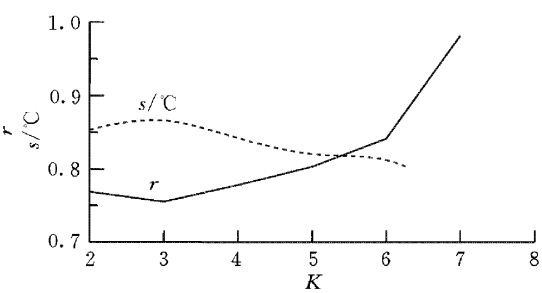

图 2　I-EOF 方案平均插补效果随截断阶数 K 的变化

Fig. 2　Mean interpolation precision as a function of truncated orders in I-EOF scheme

由表 2 可见, 随着缺测站点的增多, 各种方法的插补精度均有所下降。所有年份的 CVR 方案插补精度基本上高于其余两种方案, 尤其在缺测站点数达到 60 %($M_2=30$) 以上时, 其优势更明显。平均而言, 当缺测点达 80 %($M_2=40$), CVR 插补场均方差分别比 PCR、I-EOF 方法低 0.14、0.20, 距平相关系数(表略) 分别高 0.12 左右。PCR 与 I-EOF 方法插补效果相差不大, 前者略好于后者, 主要只在插补场数值大小上, 前者均方误差比后者更小。

缺测场呈 PD 分布型时(表 3), 在相同的缺测站点数下, 插补效果基本上仍以 CVR 法为最优。当缺测站数多于 60 %($M_2=30$) 时, PCR、I-EOF 方法的插补精度明显偏低, 平均均方误差达到 1 ℃左右, 距平相关系数(表略) 低于 0.45, 且插补精度的年际差异显著增大。图 3 给出了 1955 年 1 月缺测场呈 PD 分布, 缺测站点达到 60 %($M_2=30$) 时, 3 种方案的插补距平场及实测场。显然, 由 CVR 法得到的插补距平场与实测距平场的分布形势相当接近, 而其他两种方案则都略有差距。

进一步对比各种方案插补精度的年际变化可以发现, 所有方法插补效果的年际变化规律

大致相同。对任何缺测情形,各类方案插补效果最好的都是1954、1957年,且插补精度受缺测场变化的影响较小;而较差的则大多是1952、1958、1960年,并在缺测站点大于60 %(M_2=30)时,精度明显下降。这一事实表明3种方案的插补效果都依赖于距平场的分布特点。对比1957年与1952年1月实测平均气温距平场的分布形势(图略)可以发现,1957年1月气温距平以大尺度分布为主,而1952年距平场分布形势较为零乱,局地因素起主要作用。结合对其余年份气温距平场的普查,可以证明对距平分布较复杂的形势,3类方案插补精度均有不同程度的下降,相对而言,CVR方法仍优于其他两类插补方法。

表 **2** 缺测场呈 **SD** 分布时不同方法插补场均方误差

Table 2 Mean square errors of different interpolation schemes for SD

序号	缺测站数	插补方法	1960	1959	1958	1957	1956	1955	1954	1953	1952	1951	平均
		CVR	0.403	0.191	0.247	0.240	0.436	0.275	0.387	0.206	0.870	0.622	0.387
1	10	PCR	0.395	0.256	0.270	0.471	0.556	0.388	0.446	0.240	0.971	0.645	0.464
		I-EOF	0.450	0.266	0.306	0.413	0.501	0.473	0.491	0.370	1.075	0.633	0.625
		CVR	0.323	0.365	0.302	0.389	0.431	0.570	0.515	0.307	0.702	0.550	0.445
2	20	PCR	0.328	0.423	0.317	0.534	0.620	0.610	0.596	0.329	1.013	0.570	0.534
		I-EOF	0.414	0.413	0.374	0.434	0.526	0.635	0.485	0.413	0.983	0.560	0.523
		CVR	0.317	0.333	0.321	0.413	0.364	0.599	0.447	0.371	0.864	0.552	0.458
3	30	PCR	0.374	0.396	0.352	0.608	0.513	0.617	0.666	0.479	0.902	0.629	0.554
		I-EOF	0.405	0.393	0.348	0.525	0.528	1.174	0.446	0.424	0.989	0.574	0.753
		CVR	0.404	0.361	0.303	0.648	0.435	0.604	0.438	0.431	0.981	0.571	0.517
4	40	PCR	0.693	0.632	0.294	0.891	0.549	0.899	0.557	0.417	0.995	0.634	0.656
		I-EOF	0.434	0.400	0.380	0.887	0.734	1.656	0.669	0.472	0.980	0.543	0.716

表 **3** 缺测场呈 **PD** 分布时不同方法插补均方误差

Table 3 Mean square errors of different interpolation schemes for PD

序号	缺测站数	插补方法	1960	1959	1958	1957	1956	1955	1954	1953	1952	1951	平均
		CVR	0.584	0.459	0.320	0.418	0.388	0.597	0.409	0.365	0.915	0.546	0.500
1	10	PCR	0.685	0.546	0.309	0.473	0.498	0.745	0.613	0.594	0.633	0.422	0.552
		I-EOF	0.689	0.659	0.313	0.611	0.489	1.103	0.525	0.599	0.850	0.505	0.635
		CVR	0.456	0.356	0.320	0.415	0.521	0.521	0.522	0.492	0.998	0.571	0.517
2	20	PCR	0.562	0.527	0.313	1.251	0.590	0.759	0.619	0.548	1.185	0.378	0.673
		I-EOF	0.594	0.513	0.364	0.619	0.697	1.506	0.615	0.598	1.135	0.558	0.718
		CVR	0.752	0.400	0.540	0.621	0.677	0.728	0.686	0.469	1.308	0.680	0.686
3	30	PCR	0.566	0.422	0.369	1.179	0.780	1.347	0.927	0.549	1.520	0.516	0.817
		I-EOF	0.578	0.472	0.346	0.737	0.652	1.895	0.534	0.521	1.274	0.624	0.763
		CVR	0.565	0.323	0.414	0.837	0.792	1.521	0.950	0.621	1.596	0.745	0.836
4	40	PCR	0.613	0.453	0.385	2.913	0.664	2.016	0.575	0.467	1.652	1.245	1.098
		I-EOF	0.665	0.536	0.710	0.793	1.464	2.318	1.032	0.539	1.671	0.634	1.035

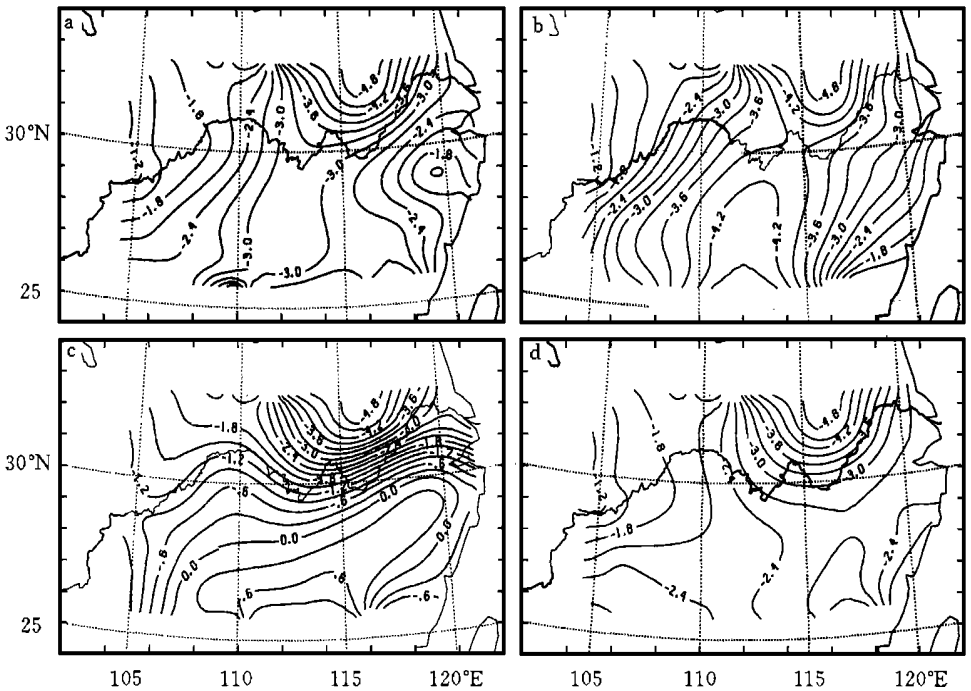

图 3　1955 年 1 月气温实测场及各种插补方案的距平插补场

a. CVR; b. PCR; c. I-EOF; d. 实测场

Fig. 3　Temperature anomaly fields observed and interpolated by different schemes in the January of 1955 (M_2= 30, under PD distribution)

a. CVR; b. PCR; c. I-EOF; d. observed field

4. 2　连续缺测多年的插补效果比较

为考察各种方案对连续缺测多年的资料场的恢复效果, 我们假设 1960～1951 年 1 月平均气温场连续缺测, 作各种方案下的插补试验, 并记由此得到的 1951～1990 年 1 月气温距平场序列为 $\hat{\boldsymbol{X}}$。为检验其插补序列的可靠性, 分别对 3 种方案得到的资料阵 $\hat{\boldsymbol{X}}$ 及实测资料阵 $\boldsymbol{X}$ 作主分量分析。

对缺测站点数 M_2= 30(达 60 %) 呈 PD 分布的缺测场作连续插补试验, 表 4 给出了它们前 4 个主分量的方差贡献及其与实测资料阵 $\boldsymbol{X}$ 主分量间的相关系数。由表可见, 各种方案前几个主分量都与实测主分量十分相似, 尤其是 CVR 方法, 其前 4 个主分量的方差贡献都只相差 1 % 以下, 前 3 个主分量与实测场主分量的相关系数也都高于 0. 85, 表明它们对大尺度分布形势的插补效果是相当理想的。

表 5 列出了缺测站点占 80 %(呈 PD 分布)时, 各种方案的主分量相似程度。显然, 由 CVR 方法得到的前 4 个主分量仍与实际值相当接近, 第 1 主分量的方差贡献只相差 1. 66 %, 前 4 个主分量的相关系数也高于 0. 85。但其他两种方法仅前 2 个主分量的相关系数在 0. 8 左右, 且由 PCR 方法得到的第 1 主分量也与实测序列差 10 %, I-EOF 法的第 1 主分量则差 7 %, 其后主分量的相关系数则都低于 0. 60。可见, 此时 CVR 方法的优越性更为明显。

表 4 不同插补方案下插补场与实测场的前 4 个主分量的比较(M_2= 30, PD 分布)

Table 4 First four principal components of observation and interpolation fields (M_2= 30, under PD distribution)

		序号				累积方差贡献
		1	2	3	4	
方差贡献(%)	实测场($\boldsymbol{X}$)	74.61	11.23	5.20	1.83	92.87
	$\hat{\boldsymbol{X}}_{\mathrm{CRV}}$	75.44	11.98	5.16	1.64	94.22
	$\hat{\boldsymbol{X}}_{\mathrm{PCR}}$	77.56	10.74	5.22	1.25	94.77
	$\hat{\boldsymbol{X}}_{\mathrm{I\text{-}EOF}}$	70.87	13.15	8.50	1.60	94.12
相关系数	$\hat{\boldsymbol{X}}_{\mathrm{CRV}}$	0.997	0.932	0.980	0.878	
	$\hat{\boldsymbol{X}}_{\mathrm{PCR}}$	0.994	0.934	0.981	0.656	
	$\hat{\boldsymbol{X}}_{\mathrm{I\text{-}EOF}}$	0.987	0.900	0.857	0.856	

表 5 不同插补方案下插补场与实测场的前 4 个主分量的比较(M_2= 40, PD 分布)

Table 5 First four principal components of observation and interpolation fields (M_2= 40, under PD distribution)

		序号				积方差贡献
		1	2	3	4	
方差贡献(%)	实测场($\boldsymbol{X}$)	74.61	11.23	5.20	1.83	92.87
	$\hat{\boldsymbol{X}}_{\mathrm{CRV}}$	76.27	11.14	5.29	1.88	94.58
	$\hat{\boldsymbol{X}}_{\mathrm{PCR}}$	84.38	7.94	2.96	1.07	96.35
	$\hat{\boldsymbol{X}}_{\mathrm{I\text{-}EOF}}$	67.45	14.79	10.67	1.56	94.49
相关系数	$\hat{\boldsymbol{X}}_{\mathrm{CRV}}$	0.990	0.942	0.940	0.852	
	$\hat{\boldsymbol{X}}_{\mathrm{PCR}}$	0.875	0.887	0.588	0.589	
	$\hat{\boldsymbol{X}}_{\mathrm{I\text{-}EOF}}$	0.945	0.757	0.547	0.409	

综上可见, 缺测站数若低于 60 %, 无论其空间分布如何, 3 种方法都有较好的插补效果, 尤以 CVR 方法精度更高。缺测站数高于 60 % 时, CVR 方法优势更明显, 且随缺测年数加长而相对另两种方法更优。

5 结 论

利用基于 EOFS 的主分量回归(PCR)、EOFS 迭代法(I-EOF) 和基于主分量典型相关的典型变量回归(CVR) 3 种不同的统计插补方案, 对同一区域同一种气象要素序列进行缺测资料的插补延长试验表明, 各种方案插补精度都与参数选择有关, 因此, 实施各方案前应作参数选择试验。

无论缺测站点空间分布类型如何, 当缺测站数较少(< 60 %) 时, 3 种方法插补效果差异不大, 且较好, 但缺测站数较多(> 60 %), 尤其是实际距平分布较为复杂时, CVR 方法优势明显, 且随缺测年数加长而更突出。可见, 主分量典型相关基础上的资料插补方法效果最好。

参 考 文 献

[1] 么枕生. 中国境内农业指标温度的出现日期、持续日期与积温[J]. 地理学报, 1957, 23(2): 183～204
[2] BROOKS C E P, Carruthers N. Handbook of statistical methods in meteorology[M]. London: Her Majesty's Stationery office, 1953
[3] 么枕生. 气候统计[M]. 北京: 气象出版社, 1963
[4] 屠其璞. 一种气温场序列插补方法[J]. 南京气象学院学报, 1986, 9(1): 19～30
[5] 江志红, 丁裕国. 我国近百年(1881～1980 年)总辐射场资料的重建试验[J]. 气象科学, 1990, 10(1): 22～30
[6] 丁裕国, 冯燕华, 袁立新. 用统计模式重建热带太平洋环流场资料的可行性试验[J]. 热带气象, 1992, 8(4): 207～305
[7] 张邦林, 丑纪范, 孙照渤. 用前期大气环流预报中国夏季降水的 EOF 迭代方案[J]. 科学通报, 1991, 36(23): 1797～1798
[8] BARNETT T P, PRESIENDORFER R. Origins and levels of monthly and seasonal forecasts kill for United States surface air temperatures determined by canonical correlation analysis[J]. Mon Wea Rev, 1987, 115(29): 1825～1850
[9] 江志红, 丁裕国. 基于PC-CCA 方法的气象场资料插补试验[J]. 南京气象学院学报, 1999, 22(2): 141～148

CONTRAST STUDY ON THE SEVERAL INTERPOLATION SCHEMES OF METEOROLOGICAL FIELDS SERIES

JIANG Zhi-hong,　DING Yu-guo,　TU Qi-pu
(Department of Environment Sciences, NIM, Nanjing　210044)

Abstract: Three different statistical interpolation schemes respectively based on the principal component regression (PCR) of empirical orthogonal function (EOF), the iteration of EOF (I-EOF) and the canonical variable regression (CVR) of principal component are adopted to interpolate the same missing data fields. Contrast experiments show that the interpolation precision of all kinds of schemes is associated with its parameters. If the pencentage of missing data station numbers is less than 60 %, no matter what distributions these stations show, all schemes, especially CVR exhibit a satisfactory accuracy, and this character becomes more significant when data-missing goes a longer time.

Keywords: meteorological field series; data series interpolation; canonical correlation analysis; empirical orthogonal function