

EOF 迭代模型的月动力延伸预报产品释用方法^{*}

池俊成 史印山

(河北省气象台, 石家庄 050021)

摘 要

以动力延伸预报产品为基础,应用 EOF 迭代模型对河北省 2005—2007 年 6—8 月逐月降水量进行预测试验。经过详细分析不同相关区因子、截取特征向量的项数、EOF 迭代次数和迭代误差对预测结果的影响,提出在建立模型时应该尽可能选取与预测对象关系密切的相关区因子。随着迭代次数增加,迭代误差逐渐减小,但预测精度与迭代误差没有必然联系。试验结果表明:迭代次数不宜过多,甚至不需要迭代,当仅取前一项特征向量,不进行迭代时,月降水量预测值和实况值的符号一致率为 0.63, PS 评分为 0.74。

关键词: EOF 迭代;月动力延伸预报产品;降水预测

引 言

动力气候模式是目前国际上开展气候预测的主要工具。经过多年研制、发展和业务化过程,国家气候中心建立了动力气候模式预测业务系统^[1-2],其预测结果在实际业务中得到应用,并逐步成为我国短期气候预测业务的重要依据。

李维京等^[3]从一组大气动力学方程组出发,推导出局地月降水距平与月平均环流场距平的关系;陈丽娟等^[4-5]在此基础上,设计了月降水降尺度模型,并在我国降水的降尺度应用中取得了较好的结果。陈丽娟等^[6]根据 500 hPa 部分环流特征量模拟效果好于整个环流形势预报,以及环流特征量和要素预报有较好相关这一事实,制作我国月平均气温预测。其他方法还有:基于天气系统特征定性判断与相似离度指数相结合制作东北地区月降水预测^[7];采用主分量和典型相关分析方法,建立 500 hPa 旬平均高度场和月平均气温、月降水量的完全预报模型,制作西南地区的气候预测^[8]。

本文应用月动力延伸预报产品建立了 EOF 迭代预测模型^[9-10],并对河北省 6—8 月逐月降水量进行了预测试验。经过详细分析高度场相关区因子、展开场截取特征向量的项数、EOF 迭代次数和迭代误差对预测结果的影响,认为该方法用于月降水量

预测是可行的。

1 资料及资料处理

资料包括:1971—2007 年河北省 121 站夏季逐月降水量;1982—2007 年 1—12 月国家气候中心 AGCM 模式北半球 500 hPa 逐候高度集合预报场,水平分辨率为 $2.5^{\circ} \times 2.5^{\circ}$ 。先利用逐候高度预报场计算出 30 d 平均高度场,从中选出 $5^{\circ} \times 5^{\circ}$ 格点资料,用 1971—2000 年平均值将降水量处理为距平百分率,并进行标准化处理。

2 方法概述

原始资料矩阵

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix} \quad (1)$$

式(1)中, \mathbf{X}_{11} , \mathbf{X}_{12} 为 500 hPa 月平均高度场资料, \mathbf{X}_{21} 为降水量资料, \mathbf{X}_{22} 为通过迭代运算得到的预测值。

资料的存放顺序为

$$\mathbf{X}_{11} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_1} \\ \cdots & \cdots & \cdots \\ h_{m_1,1} & \cdots & h_{m_1,n_1} \end{bmatrix},$$
$$\mathbf{X}_{12} = [h_{1,n_1+1} \quad \cdots \quad h_{m_1,n_1+1}]^T$$

* 北京区域气象中心科技创新基金项目(BRMCCJ200702)资助。
2008-01-22 收到,2008-08-08 收到再改稿。

其中, m_1 为高度场资料空间格点数, $n_1 + 1$ 为高度场资料时间序列长度, T 表示矩阵的转置。

$$\mathbf{X}_{21} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n_1} \\ \cdots & \cdots & \cdots \\ r_{m_2,1} & \cdots & r_{m_2,n_1} \end{bmatrix}$$

其中, m_2 为降水测站数, n_1 为降水量资料时间序列长度。

令初始场 $\mathbf{X}_{22} = 0$, 构造

$$\mathbf{X}^{(0)} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & 0 \end{bmatrix} \quad (2)$$

对矩阵 $\mathbf{X}^{(0)}$ 进行 EOF 展开, 将特征值按升序排列, 截取前 K 项对应的特征向量得到拟合场

$$\hat{\mathbf{X}}^{(0)} = \mathbf{T}^{(0)} \mathbf{V}^{(0)} = \begin{bmatrix} \hat{\mathbf{X}}_{11}^{(0)} & \hat{\mathbf{X}}_{12}^{(0)} \\ \hat{\mathbf{X}}_{21}^{(0)} & \hat{\mathbf{X}}_{22}^{(0)} \end{bmatrix} \quad (3)$$

构造矩阵

$$\mathbf{X}^{(1)} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \hat{\mathbf{X}}_{22}^{(0)} \end{bmatrix} \quad (4)$$

给定迭代误差 ϵ , 重复作 EOF 展开, 直至 $\|\hat{\mathbf{X}}_{22}^{(i)} - \hat{\mathbf{X}}_{22}^{(i-1)}\| < \epsilon$ 时停止迭代运算, 终值 $\hat{\mathbf{X}}_{22}^{(i)}$ 即为预测值。

要素场和高度场的空间点总数远大于资料时间序列长度, 由于 $\mathbf{X}\mathbf{X}^T$ 与 $\mathbf{X}^T\mathbf{X}$ 的秩相同, 既非零的特征值相同, 可以利用时空转换方法^[11]求解特征值和相应的特征向量。

3 预测试验

预测的范围为河北省 121 个测站, 时间为 2005—2007 年 6—8 月, 共计 9 个月降水量。试验结果采用符号一致率 $\rho_s = (\rho_+ + \rho_-) / L$ 进行评估, 其中 ρ_+ 与 ρ_- 分别表示预测值与实况值同为“+”号或“-”号的测站数, $L = 121$ 为测站总数。

3.1 取不同相关区的对比试验

先将高度场处理成纬偏场^[12], 以便突出环流异常的驻波性质

$$h'_{i,j} = h_{i,j} - \frac{1}{m} \sum_{j=1}^m h_{i,j} \quad (i = 1, \dots, n; j = 1, \dots, m) \quad (5)$$

式(5)中, n 为沿经圈资料的格点数, m 为沿纬圈资料的格点数, 计算某站月降水量与纬偏场各点的相关系数 $r_{i,j}$, 得到相关系数场。

在相关系数场中, 给定相关系数的临界值 r_c , 当 $|r_{i,j}| > r_c$ 时, $r_{i,j} = 1$, 否则 $r_{i,j} = 0$, 由此得到 0~1 分布场。

将河北省 121 个站的 0~1 分布场对应相加, 得到相关系数绝对值大于 r_c 的频数分布场, 选取频数大于某一数值的区域即为相关区。

为了考察选取不同相关区因子对预测结果的影响, 选取 6 个不同相关系数临界值对应的相关区和全部 $5^\circ \times 5^\circ$ 格点资料。当模型截取前 4 项特征向量, 迭代误差 $\epsilon = 0.05$ 时, 试验结果见表 1。可以看出, 就河北省夏季降水而言, 临界相关系数在 0.45~0.50 之间, 相关区内格点数在 45~90 范围内预测效果最好, 符号一致率平均为 0.54, 标准差也达到最小, 表明预测结果相对稳定。

相关区因子的选取对预测效果影响的检验。 u_i 为使用全部 1296 个格点资料预测评估结果, v_i 为相关系数临界值 $r_c = 0.45$, 使用 90 个点相关区因子时的预测评估结果, 如果两组成对数据的差值 $x_i = u_i - v_i$ ($i = 1, 2, \dots, 9$), 可以看作来自正态分布 $N(0, \sigma^2)$ 的样本, 则统计量

$$T = \frac{\bar{x}}{\sqrt{S^2/9}} = 3.5581 \quad (6)$$

式(6)中, $S^2 = \frac{1}{8} \sum_{i=1}^9 (x_i - \bar{x})^2$, 服从 8 个自由度的 t 分布。当显著水平 $\alpha = 0.01$ 时, $P\{|T| > 3.3554\} =$

表 1 选取不同相关系数临界值的符号一致率
Table 1 Sign coincidence rate in different critical correlation coefficient

相关系数 临界值	相关区内 格点数	2005 年			2006 年			2007 年			平均值	标准差
		6 月	7 月	8 月	6 月	7 月	8 月	6 月	7 月	8 月		
0	1296	0.37	0.26	0.45	0.48	0.38	0.67	0.46	0.43	0.27	0.419	0.123
0.15	610	0.39	0.23	0.38	0.53	0.61	0.64	0.47	0.79	0.15	0.466	0.204
0.35	310	0.39	0.52	0.40	0.57	0.55	0.72	0.57	0.63	0.17	0.502	0.162
0.35	150	0.37	0.45	0.37	0.60	0.55	0.67	0.55	0.77	0.29	0.513	0.157
0.45	90	0.37	0.50	0.47	0.57	0.50	0.69	0.60	0.72	0.47	0.543	0.113
0.50	45	0.37	0.42	0.48	0.53	0.57	0.59	0.60	0.65	0.65	0.540	0.099
0.50	25	0.45	0.47	0.33	0.57	0.52	0.58	0.58	0.74	0.50	0.527	0.113

0.01,说明在建立预测模型时,选取与预测对象关系密切的相关因子对预测质量的提高显著。

3.2 截取不同特征向量项数对预测结果影响

当相关系数临界值 $r_c=0.45$,相关区内格点数为90,迭代误差 $\epsilon=0.05$ 时,令截取特征向量的项数 $K=1,2,\dots,11$,分别对各月降水量进行预测试验。 $K=1$ 时,累积方差贡献约为40%; $K=2$ 时,累积方差贡献约为50%,当 $K=11$ 时,累积方差贡献

达到84%。 $K=1$ 或 $K=2$ 时预测效果最好,9个月预测水平的平均符号一致率分别为0.60,0.58(图1a),这一结果与文献[13-14]得出的结论基本相同。随着 K 的增加,预测效果有所下降,但下降的幅度不大,平均符号一致率基本在0.54~0.52区间内。值得注意的是 $K=1,2$ 时,平均符号一致率最高,但方差也最大(图1b),因此,当取截取特征向量的项数较少时,应该注意预测结果的稳定性。

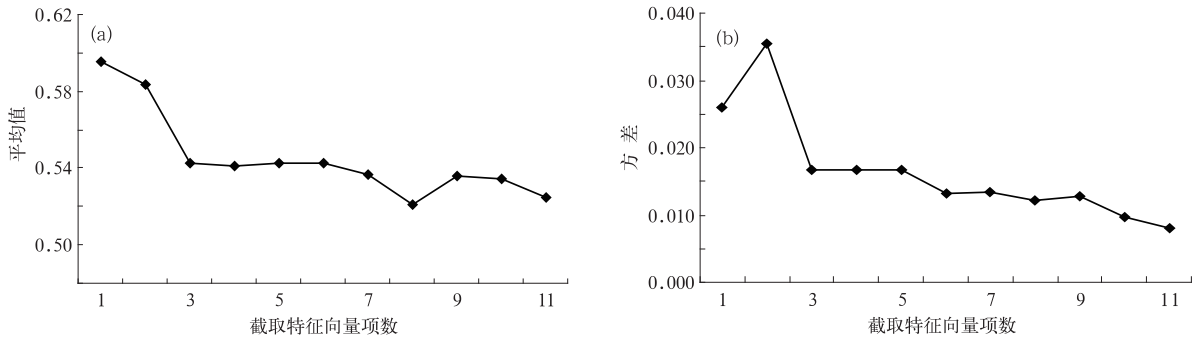


图1 符号一致率平均值(a)、方差(b)随截取特征向量项数的变化

Fig. 1 Mean(a) and variance(b) of sign coincidence rate as a function of truncation orders

3.3 迭代次数对预测结果影响

当相关系数临界值 $r_c=0.45$,相关区内格点数为90,截取前4项特征向量,令迭代次数 $S=0,\dots,19$ 分别进行预测试验。迭代次数增加时,迭代误差 ϵ 逐渐减小,即迭代运算是收敛的,但不同预测对象收敛速度不同,同样 $\epsilon=0.05$ 时,最快仅需要1次迭代运算即可达到给定的精度,最慢则需要迭代15次。计算结果表明,迭代次数与预测结果两者之间并不存在确定关系,如2007年7月预测试验中,随着迭代次数增加,预测质量逐渐提高;2007年8月预测试验中,随着迭代次数增加,预测质量反而下降;2005年7月预测试验中,随着迭代次数增加,预测质量变化较为平稳。图2给出预测质量随迭代次数变化3种不同情况。计算符号一致率与迭代误差之间的相关系数,上述3种情况的相关系数分别为-0.44,0.86,0.12。前两个相关系数分别通过了 $\alpha=0.05$ 和 $\alpha=0.001$ 的显著性检验。这一现象说明随着迭代次数增加, ϵ 减小,但预测效果不一定会得到改善。

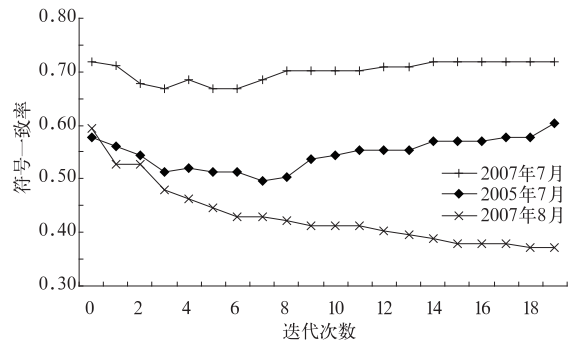


图2 符号一致率与迭代次数的关系

Fig. 2 Relationship between sign coincidence rate and iteration number

表2 最佳预测结果出现的迭代次数、符号一致率、迭代误差

Table 2 Iteration number, sign coincidence rate and iteration error in best prediction

时间	迭代次数	符号一致率	迭代误差
2005-06	0	0.413	0.389
2005-07	19	0.603	0.036
2005-08	0	0.504	0.294
2006-06	2	0.595	0.087
2006-07	0	0.578	0.313
2006-08	1	0.694	0.141
2007-06	0	0.603	0.384
2007-07	0	0.719	0.161
2007-08	0	0.595	0.320

表2 给出 $r_c=0.45$ 时,达到最佳预测结果所需的迭代次数和迭代误差。由表2可见,除2005年7月的预测最佳效果出现在迭代19次以后(实际出现在27次,符号一致率为0.652),其他情况达到最佳

预测效果一般不超过 2 次, 大多数的情况下甚至不需要迭代。

当 $r_c = 0.45$, 仅取前一项特征向量, 不进行迭代时, 月降水量预测值和实况值的符号一致率平均为 0.63 (表 3)。为了便于比较, 表 3 中还给出业务中使用的预测评分 (PS) 和技巧评分 (SS)^[15]。

表 3 2005—2007 年 6—8 月降水预测评分
Table 3 Score of precipitation from June to August during 2005—2007

时间	符号一致率	预测评分(PS)	技巧评分(SS)
2005-06	0.62	0.71	0.48
2005-07	0.67	0.83	0.69
2005-08	0.54	0.74	0.53
2006-06	0.45	0.58	0.19
2006-07	0.66	0.75	0.56
2006-08	0.65	0.74	0.48
2007-06	0.45	0.59	0.14
2007-07	0.75	0.85	0.73
2007-08	0.84	0.90	0.82

4 小 结

EOF 迭代模型与其他统计方法一样, 预测效果的好坏与选取的预测因子有关, 因此在建立预测模型时, 应该选取与预测对象关系密切的相关区因子。

相关系数临界值 $r_c = 0.45$, 展开场截取特征向量的项数 $K = 1$ 或 $K = 2$ 时, 累积方差贡献约为 40%~50%, 预测值符号一致率平均分别为 0.60, 0.58, 随着 K 的增加, 预测效果有所下降。

随着迭代次数增加, 迭代误差逐渐减小, 即迭代运算是收敛的, 但预测精度与迭代误差没有必然联

系。试验结果表明, 迭代次数对预测结果有一定影响, 但影响存在不确定性。

致 谢: 国家气候中心李维京研究员对本文提出许多宝贵意见和建议, 在此表示衷心的感谢!

参 考 文 献

- [1] 李维京, 张培群, 李清泉. 等. 动力气候模式预测系统业务化及其应用. 应用气象学报, 2005, 16(增刊): 1-10.
- [2] 丁一汇, 李清泉, 李维京, 等. 中国业务动力季节预报进展. 气象学报, 2004, 62(5): 598-612.
- [3] 李维京, 陈丽娟. 动力延伸预报产品释用方法的研究. 气象学报, 1999, 57(3): 338-344.
- [4] 陈丽娟, 李维京, 张培群, 等. 降尺度技术在月降水预报中的应用. 应用气象学报, 2003, 14(6): 648-655.
- [5] 江双武, 田红, 陈丽娟. 动力延伸预报产品释用方法的改进试验. 应用气象学报, 2005, 16(6): 779-786.
- [6] 陈丽娟, 李维京. 月动力延伸预报产品的评估和解释应用. 应用气象学报, 1999, 10(4): 486-490.
- [7] 张立祥, 陈力强, 刘文明, 等. 东北地区夏季月降水数值产品释用预报方法. 应用气象学报, 2000, 11(3): 348-354.
- [8] 马振锋, 陈洪. T63 月延伸预报在西南区域短期气候预测中的应用研究. 应用气象学报, 1999, 10(3): 368-373.
- [9] 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. 科学通报, 1991, 36(23): 1797-1798.
- [10] 李跃清, 董文林. 用多时次资料的 EOF 迭代温度长期预报. 气象, 1995, 21(9): 30-33.
- [11] 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 2000.
- [12] 李栋梁, 姚辉. 中国西北夏季降水量与 500 hPa 纬偏场的特征分析. 气象, 1995, 21(11): 22-26.
- [13] 段旭, 尤卫红, 李跃清. 多时次资料的 EOF 迭代在云南夏季气候预测中的应用. 高原气象, 2001, 20(2): 220-223.
- [14] 张永领, 吴胜安, 丁裕国, 等. SVD 迭代模型在夏季降水预测中的应用. 气象学报, 2006, 64(1): 121-127.
- [15] 陈桂英, 赵振国, 王绍武. 短期气候预测评估方法和业务初估. 应用气象学报, 1998, 9(2): 178-185.

Application Method of Monthly Dynamics Extended-range Forecast Based on EOF Iteration Scheme

Chi Juncheng Shi Yinshan

(Hebei Provincial Meteorological Observatory, Shijiazhuang 050021)

Abstract

Based on EOF iteration scheme model with dynamic extended-range forecast products, experiment of precipitation prediction is done for 121 station of Hebei Province from June to August in 2005—2007. It is necessary to use 500 hPa correlation area in the EOF iteration scheme model in order to improve the prediction accuracy. Experiment in different correlation areas indicates that the model performs best when the critical correlation coefficient is 0.45, correlation area is of 90 grids. The average sign coincidence rate between predicted precipitation and observed precipitation is 0.54.

When critical correlation coefficient is 0.45, grid number is 90 and iteration error is 0.05, let truncation order of eigenvector equal 1, 2, ..., 11 and do experiment separately. The first two eigenvectors account for about 50% of total variance explained. When truncation order equal 1 and 2, average sign coincidence rate are 0.60 and 0.58. When truncation order is greater than 2, average sign coincidence rate decreases as truncation order increases slightly until stability in 0.54—0.52.

When critical correlation coefficient is 0.45, grid number is 90 and truncation order of eigenvector is 4, let iteration number equal 0, 1, ..., 19 do the prediction experiment separately. The experiment shows that iteration error decreases as iteration number increases, but the decrease rate is very different. As accuracy reaches 0.05, the fastest experiment requires only one times of iterative operator and the slowest one requires 15 times. The predicted result affected by iteration error is slight. To acquire good result, many times of iteration is unnecessary, usually no more than three times. When truncating the first eigenvector without iteration, the average sign coincidence rate of monthly precipitation prediction can reach 0.63 and operational scores PS can reach 0.74.

Key words: EOF iteration; monthly dynamics extended-range forecast; precipitation prediction