提高月降水量预报精度的一种尝试

张永革　　王延贵

（山东省枣庄市气象局，277148）　　（山东济宁市气象局）

提　　要

用原始数据建立的多元自回归预报模型，虽然对气压，气温、绝对湿度的预报精度很高，但对月降水量的预报精度较低。改用自典型相关因子建模，大大提高了月降水量的预报精度，对气压、气温、绝对湿度的预报亦有所改进。

关键词：自典型相关因子　月降水量　预报精度

引　　言

在《时间序列分析与动态数据建模》(1987 年版)一书中，有这样一个问题：根据 334 页上表 1 中的原始数据，建立了 4 个变量三阶多元自回归预报模型：

$$x_t = \hat{\varphi}_0 + \hat{\varphi}_{31}x_{t-1} + \hat{\varphi}_{32}x_{t-2} + \hat{\varphi}_{33}x_{t-3} + \varepsilon_t \tag{1}$$

我们据此计算了月降水量、气压、气温、绝对湿度的预报误差方差。从 117 组数据的预报误差来看，此模型对气压、气温、绝对湿度的预报精度很高，而月降水量的预报精度较低。为此我们试图用一种新的建模方法，以提高月降水量的预报精度。

1　自典型相关分析模型

多元统计分析中典型相关分析的基本思路是：分别在两组随机变量中选取若干有代表性的综合指标，也就是所谓典型变量，通过这两组典型变量之间的相关关系的研究，代替对原来两组原始变量的相关关系的研究。典型相关分析应用于时间序列中建立的数学模型，称为自典型相关分析模型。根据该书 336 页表 2 中样本相关阵的特点分析，我们认为气压、气温、绝对湿度之间可能存在很强的自典型相关因子，故我们期望通过建立自典型相关分析数学模型来达到我们的目的。

对自典型相关分析的两组变量定义如下：

对于时间序列中任一时刻 t，构造多元数据的下列两个线性组合，作为自典型相关分析的两组典型相关变量：

$$U_t = \sum_{i=1}^{p} l_i x_{it} \tag{2}$$

$$V_t = \sum_{k=1}^{s}\sum_{i=1}^{p} m_{i,t-k}x_{i,t-k} \tag{3}$$

其中 p 为多元数据变量数，s 为所考虑的滞后最大值。

若记 $\boldsymbol{L} = (l_1, l_2, \cdots\cdots l_p)^T$

$$\boldsymbol{M} = (m_{1,t-1}, m_{2,t-1}, \cdots\cdots, m_{p,t-1}, m_{1,t-2}, m_{2,t-2}, \cdots\cdots, m_{p,t-2}, \cdots\cdots, m_{1,t-s}, \cdots\cdots m_{p,t-s})^T$$

$$\boldsymbol{X}_t = (x_{1t}, x_{2t}, \cdots\cdots, x_{pt})^T$$

$$\boldsymbol{Y}_t = (x_{1,t-1}, x_{2,t-1}, \cdots\cdots, x_{p,t-1}, x_{1,t-2}, x_{2,t-2}, \cdots\cdots, x_{p,t-2}, \cdots\cdots, x_{1,t-s}, x_{2,t-s}, \cdots\cdots, x_{p,t-s})^T$$

$\boldsymbol{L}$ 为 p 维列向量，M 为 $p\times s$ 维列向量，X_t 为 p 维列向量，Y_t 为 $p\times s$ 维列向量。

设 $\boldsymbol{X}_t$ 的协方差矩阵为 $\boldsymbol{D}_{11}$，$\boldsymbol{Y}_t$ 的协方差矩阵为 $\boldsymbol{D}_{22}$，$\boldsymbol{X}_t$ 与 $\boldsymbol{Y}_t$ 的协方差矩阵为 $\boldsymbol{D}_{12}$，$\boldsymbol{Y}_t$ 与 $\boldsymbol{X}_t$ 的协方差矩阵为 $\boldsymbol{D}_{21}$，$\boldsymbol{X}_t$ 与 $\boldsymbol{X}_{t-k}$ 的协方差矩阵为 $\boldsymbol{D}_{11}^{(k)}(k=1,2,\cdots\cdots,s)$。在平稳性的条件下，$\boldsymbol{D}_{22}$、$\boldsymbol{D}_{12}$、$\boldsymbol{D}_{21}$ 可分别表示如下：

$$\boldsymbol{D}_{22} = \begin{pmatrix} D_{11} & D_{11}^{(1)} & D_{11}^{(2)} & \cdots & D_{11}^{(s-1)} \\ D_{11}^{(1)} & D_{11} & D_{11}^{(1)} & \cdots & D_{11}^{(s-2)} \\ \vdots & \vdots & \vdots & & \vdots \\ D_{11}^{(s-1)} & D_{11}^{(s-2)} & D_{11}^{(s-3)} & \cdots & D_{11} \end{pmatrix} \tag{4}$$

因为 $\boldsymbol{D}_{11}$、$\boldsymbol{D}_{11}^{k}(k=1,2,\cdots\cdots s)$ 均为 p 阶

对称矩阵，所以 $\boldsymbol{D}_{22}$ 为 $p\times s$ 阶对称矩阵。

$$\boldsymbol{D}_{12}=(D_{11}^{(1)},D_{11}^{(2)},\cdots\cdots D_{11}^{(s)}) \quad (5)$$

根据 $\boldsymbol{D}_{11}^{(k)}$ 的对称性，则有：

$$\boldsymbol{D}_{21}=(\boldsymbol{D}_{11}^{(1)},\boldsymbol{D}_{11}^{(2)},\cdots\cdots,\boldsymbol{D}_{11}^{(s)})^{\mathrm{T}}=D_{12}^{\mathrm{T}} \quad (6)$$

因为 $\boldsymbol{D}_{12}$ 是 p 行 $p\times s$ 列矩阵，所以 $\boldsymbol{D}_{21}$ 是 p×s 行 p 列矩阵。

令 $\boldsymbol{A}=\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{21}=\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{12}^{\mathrm{T}} \quad (7)$

$$\boldsymbol{B}=\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{21}\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12}=\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{12}^{\mathrm{T}}\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12} \quad (8)$$

根据典型相关分析理论可得：

$$AL=\lambda^2 L \quad (9)$$

$$BM=\lambda^2 M \quad (10)$$

λ 恰为两组线性组合 U_t 与 V_t 之间的简单相关系数，为使其达到最大，可由 A 中求出最大特征值 λ^2（由 B 中求出亦可），据此所求 $\boldsymbol{L}$、$\boldsymbol{M}$ 即分别为 A 和 B 的最大特征值所对应的特征向量。p 个变量应有 p 个特征值，若特征值按由大到小的顺序排列，所对应的 p 特征向量 L_i 与 $M_i(i=1,2,\cdots\cdots,p)$ 称为 p 对自典型相关因子。它们的自典型相关系数，也按由大到小的顺序排列。

为了与模型(1)进行比较，我们也将最大滞后 s 取为 3。

根据文献[1]334 页表 1 中气压、气温、绝对湿度三个变量的 120 组数据，按(9)、(10)两式分别求出矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$。对非对称矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 应用 QR 算法求出它们的全部特征值和特征向量。今将所求三个特征值及自典型相关系数和全部特征值对应的三对自典型相关因子列于表 1。

表 1

序号	特征值 λ_i^2	自典型相关系数 λ_i	自典型相关因子向量（特征向量）L_i，M_i
1	0.9240	0.9612	$\boldsymbol{L}_1$=(−0.5320,0.5522,0.7756) $\boldsymbol{M}_1$=(−0.3082,0.6967,0.2941,−0.0996,0.2164,−0.3078,0.2753,−0.2096,−0.1589)
2	0.4949	0.7035	$\boldsymbol{L}_2$=(−0.6814,−0.7293,0.0604) $\boldsymbol{M}_2$=(−0.4055,−0.2809,−0.1507,−0.1039,0.1008,−0.1543,0.2352,0.2842,−0.0874)
3	0.3326	0.5767	$\boldsymbol{L}_3$=(−0.1995,0.5820,−0.7741) $\boldsymbol{M}_3$=(0.3611,−0.3311,0.8028,0.2024,0.5300,−0.5960,−0.1860,0.9078,−0.9528)

由此可见，气压、气温、绝对湿度确实存在很强的自典型相关因子。

由于我们的工作是在满足平稳性的假设下进行的，因此对所求的自典型相关系数需要检验。根据所求的特征向量对 117 组（$N-S=120-3=117$）样本计算 U_t 和 V_t，然后计算出 U_t 和 V_t 的相关系数，其值为 0.9681，而本例最大自典型相关系数则为 0.9612，误差很小，属于计算误差，因此可认为平稳性条件得到满足，所求自典型相关因子有意义。

2 月降水量预报

设 x_{1t} 代表 t 时刻月降水量，x_{2t}、x_{3t}、x_{4t} 分别代表 t 时刻月平均气压、气温、绝对湿度。

月降水量预报方程为：

$$\begin{aligned}x_{1t}=&c_0+c_1\boldsymbol{Y}_t^{\mathrm{T}}\boldsymbol{M}_1+c_2\boldsymbol{Y}_t^{\mathrm{T}}\boldsymbol{M}_2+c_3\boldsymbol{Y}_t^{\mathrm{T}}\boldsymbol{M}_3+\\&c_4\boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{L}_1+c_5\boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{L}_2+c_6\boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{L}_3+\\&c_7\boldsymbol{X}_{t-2}^{\mathrm{T}}\boldsymbol{L}_1+c_8\boldsymbol{X}_{t-2}^{\mathrm{T}}\mathbf{L}_2+c_9\boldsymbol{X}_{t-2}^{\mathrm{T}}\boldsymbol{L}_3+\\&c_{10}\boldsymbol{X}_{t-3}^{\mathrm{T}}\boldsymbol{L}_1+c_{11}\boldsymbol{X}_{t-3}^{\mathrm{T}}\boldsymbol{L}_2+c_{12}\boldsymbol{X}_{t-3}^{\mathrm{T}}\boldsymbol{L}_3+\\&c_{13}\boldsymbol{X}_{1,t-1}+c_{14}\boldsymbol{X}_{1,t-2}+c_{15}\boldsymbol{X}_{1,t-3}\end{aligned} \quad (11)$$

其中：

$$\boldsymbol{X}_{t-k}=(x_{2,t-k},x_{3,t-k},x_{4,t-k})^T \quad k=1,2,3$$

$$\boldsymbol{Y}_t=(x_{2,t-1},x_{3,t-1},x_{4,t-1},x_{2,t-2},x_{3,t-2},x_{4,t-2},x_{2,t-3},x_{3,t-3},x_{4,t-3})^T$$

系数向量 $\boldsymbol{c}=(c_0,c_1,c_2,\cdots,c_{15})$

利用最小二乘法求得系数向量 $\boldsymbol{c}$ 为：

$\boldsymbol{c}$ = (87.32, 79.5359, 121.3007, 72.8930,

−56.1144， − 38.1464， 23.0251，
7.4963，
12.7948，−106.1085，21.2921，
66.8249，−89.8862，0.0201，
−0.1004，0.0391）

对117组数据（$N-S=120-3=117$）计算两模型的误差方差，为便于比较，将其列于表2中。

表2

预报对象	误差方差	
	用原始变量建模	用自典型相关因子建模
降水量	3009.2950	2870.8060
气压	3.5908	3.4789
气温	3.5963	3.4871
绝对湿度	4.5289	4.4172

我们又根据枣庄市的120组数据用两种方法建模，并将最大滞后也定为3，对1996年6月份降水量、气压、气温、绝对湿度进行预报，预报结果见表3。

表3

预报对象	用原始变量建模	用自典型相关因子建模	观测值
降水量/mm	90.2	87.1	84.8
气压/hPa	996.3	996.8	996.9
气温/℃	25.7	25.4	25.5
绝对湿度/(g·m^{-3})	21.2	20.9	21.0

据此可知，用自典型相关因子建模对4个气象变量的预报效果均比用原始变量建模的效果好，其中降水量尤为突出。虽然两者都是用前3个月（$s=3$）的降水量、气压、气温和绝对湿度的12个数据来预报下一个月的降水量，但用自典型相关因子建模时，用的是原始变量气压、气温和绝对湿度的三对典型相关因子加上降水量原始变量的滞后数据，是寻求预报对象以外的变量组合的典型相关因子，自典型相关因子中均包含了降水量变量，而降水量变量与其它变量及自身滞后相关均较弱，故不如气压、气温、绝对湿度三变量组合的自典型相关因子典型，自典型相关因子更能反映气象演化过程的趋势。正因为对降水量的预报使用了最能反映气象演化过程趋势的气压、气温、绝对湿度三变量组成的自典型相关因子，所以预报精度明显提高。

3 结论

用自典型相关因子建模较之用原始变量建模预报效果好得多，不仅大大提高了月平均降水量的预报精度，并且对气压、气温、绝对湿度的预报也有明显的改进。

参考文献

1 扬位钦，顾岚编著. 时间序列分析与动态数据建模. 北京工学院出版社，1987.
2 吴国富等编. 实用数据分析方法. 北京：中国统计出版社，1992.
3 周光亚等编. 多元统计分析. 北京：地质出版社，1982.

An Attempt to Improve Accuracy in Monthly Rainfall Forecast

Zhang Yongge
(Zaozhuang Meteorological Office, Shandong 277148)

Abstract

Although the plural autogression forecast model based on primary data can get high accuracy in atmospheric pressure, temprature and absolute humidity forecasts, its application to monthly rainfall forecast is unsatisfactory. The model based on typical factor of autocorrelation was suggested, which can get much higher accuracy in monthly rainfall forecast while improves the accuracy in atmospheric pressure, temperature and absolute humidity forecasts as well.

Key Words: typical factor of autocorrelation monthly rainfall forecast accuracy