

多变量状态空间预报法在旬雨量预报中的应用*

严中伟

(中国科学院大气物理研究所, 北京 100029)

提 要

结合单个时间序列构造状态空间进行预报和传统研究中多要素分析的思路, 建立一种客观的综合多要素的状态空间预报法, 应用于旬雨量预报. 结果表明: 考虑雨量和气温的多变量状态空间比仅考虑雨量序列具有更好的预报性能——437对值的预报相关提高5%, 显示了多变量状态空间预报的可行性和潜力.

关键词: 多变量; 状态空间; 旬雨量变差; 预报.

引 言

近年来, 随着非线性系统理论在气象界的广泛应用^[1-4], 人们开始重新审视以往基于线性理论认识的统计预报方法. 我们知道, 气候跃变现象有助于理解线性方法的缺陷. 研究表明, 在很多尺度上都存在气候跃变^[5, 6], 说明任一时段的气候统计特征可能随时间很快改变, 因而对气候序列作再精确的统计拟合, 也难以预测未来. 例如用主周期分量可很好地拟合时间序列, 却难以据此预测气候^[7]. 从统计学角度来看, 气候序列是高度非平稳的, 基于平稳时间序列理论的预报方法不可能获得好结果.

因此, 一些基于非线性理论构造的预报法应运而生. 例如: Cuomo等研制了局部自回归法^[8], 并用于单站气温预报; Sugihara和May从混沌力学研究中发展的相空间单体映射法^[9, 10]在预报南方涛动指数变化的应用中效果良好; Lin等借用非线性物理研究中的概念提出长期预报的自组织法^[11]. 这些新方法共同点是: 首先确定预报起点的状态特征, 它可由状态空间中一个点来表示; 选择其相近态, 即状态空间中一系列相近点; 再由后者的局部演变规律作预报. 状态空间法更好地反映了气候序列的非平稳性, 且隐含了非线性系统动力学规律, 具有更好的预报性能. 出于理论上的考虑, 多数研究是针对单个时间序列而作的. 气象界则强调多要素分析. 结合这两条研究思路中的优势, 建立一种多变量状态空间预报法, 以改善当前气候预报(特别是降水预报)的水平, 这是本文的出发点.

Van den Dool在总结美国各州用环流型预报月气候距平时发现: 气温预报相关平均

* 国家科委 PD85-27 和自然科学基金资助, 同时得到大气科学和地球流体动力学数值模拟国家重点实验室(LASG)支持.

1995-07-03 收到, 1996-02-28 收到修改稿.

可达40%左右,一些地区甚至达80%;而降水预报相关仅维持在10%左右^[12].这反映了降水预报的难度.本文将就我国汛期预报的难题——旬雨量预报问题提出一种新的方法,并探讨其应用潜力.

1 资 料

基本资料是采用中国气象局整编的1951~1990年336站旬降水和温度资料集.单站降水观测中的局地性是降水预报的困难因素之一.即使一个地区有降水云系发展,该区内某些站点仍可能测不到雨量.因此单站记录不足以反映一个区域的降水状况,区域化处理是必要的.

选取资料集中的自1951年第一旬开始就有记录的台站,这些站缺测很少.在长江下游和海、黄河下游两个区内分别选取相关达50%以上的台站,求得区域平均的旬雨量序列.温度资料类似处理.

受季风控制的区域的雨量变化具有突出的年循环特征,即冬季少雨夏季多雨.功率谱分析表明:相对于年循环,其它尺度变化可视为“噪音”.然而对于气候预报而言,恰恰是这些“噪音”,也即相对于气候平均年循环的雨量偏差是重要的.为此我们扣除常年平均的年循环得到距平序列,把旬雨量距平序列相邻点两两相减可得变差序列.只要知道某旬的雨量距平及其变差,就可求得下旬的雨量距平.从这个意义看,预报变差和预报雨量距平没有本质差别.但从构造状态空间的效果看,变差序列在状态空间中形成的点集较为集中^[9],有利于减小预报模型的不稳定性.因此,作为对一种方法的效果检验,本文将针对变差序列来构造预报模型.

图1显示了华东区旬雨量变差序列的概貌.和温度、气压等较缓变的要素相比,这种有关雨量的序列无持续性可言.这也是降水预报水平偏低的原因之一.

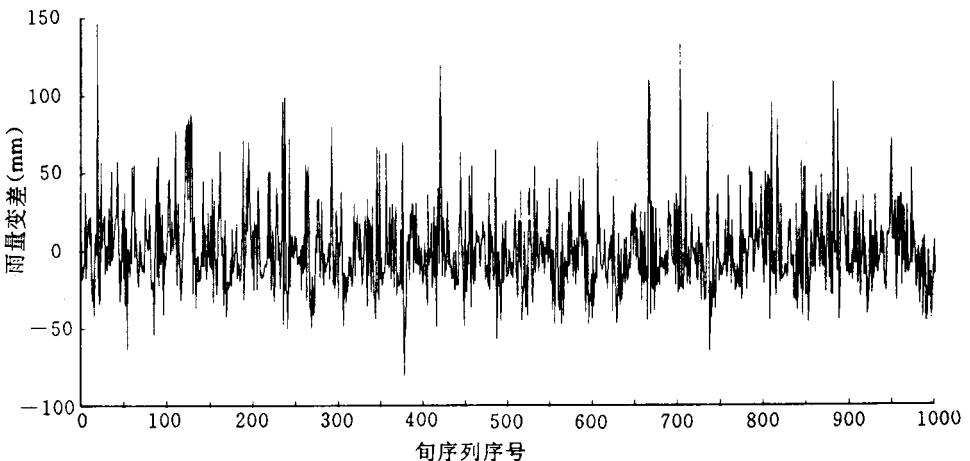


图1 华东区旬雨量序列前1000个旬的雨量变差(单位: mm)

Fig. 1 The first 1000 variable of dekad rainfall series in East China (unit: mm)

2 预报原理

设有时间序列 $X_i, i=1, 2, \dots, N$. 取嵌入空间维数 E , 滞后参数 τ , 则序列的任一下片断 $P_k = \{X_k, X_{k+\tau} \dots X_{k+(E-1)\tau}\}$ 可视为 E 维空间中一个点. 顺序取 k , P_k 就构成状态空间中随时间演变的轨迹. 理论上已证明: 适当选取 E 和 τ , 上述轨迹就可近似反映原系统动力学规律. 如果系统不是纯随机的, 且观测序列足够长而精确, 则状态空间中较接近的两点的轨迹走向在有限时间内应近于平行. 这样就可建立预报模型:

$$P'_0 = F(P_0, P_1, P_2, \dots, P_m, P'_1, P'_2, \dots, P'_m, \Delta t)$$

其中 P_0 为预报起点, P'_0 为经 Δt 时间后的预报终点, $P_1, P_2 \dots P_m$ 为状态空间中最接近 P_0 的 m 个点, 它们经 Δt 时间后变到 $P'_1, P'_2 \dots P'_m$. 这一关系的实质内容是: 预报终点取决于预报起点附近状态的变化规律. 这是状态空间法与那些基于所有观测的统计平均特征而构造的预报方法的本质差别.

最简明的一种 F 形式是: 取 P_0 的最近点 ($m=1$), 令 $\overrightarrow{P_0 P'_0}$ 平行于 $\overrightarrow{P_1 P'_1}$ 且等长. 但其应用效果不一定好. 因为实际序列被嵌入状态空间时会有许多点接近 P_0 , 其中某点成为最近点完全有可能由误差等随机因素造成. 因此本文采用如下规则: 先在 P_0 附近取半径为 r 之球域 S_0 , 设 S_0 内的点 P_1, P_2, \dots, P_m 经 Δt 时间变到 P'_1, P'_2, \dots, P'_m 构成集合 S'_0 , 令 P'_0 为该点集的平均坐标点 (或称中点). 参数 r 应足够大以确保 S_0 内有一定数量的状态点, 又要足够小以保障各点为 P_0 之相邻点, 可根据序列之变率大小和观测精度来决定. 计算程序则以获取最佳预报效果为标准来自动调整其值.

考虑多个变量的时间序列. 如果它们属于同一动力学系统, 则由其各自重建的相空间轨迹演变特征在拓扑意义上应是一致的. 另一方面, 以这些变量本身为坐标构成的空间就是原系统相空间的一个子空间, 从而使变量组的序列点到该空间所形成的演变规律与原系统规律不会矛盾. 这样, 同时考虑多个变量序列所确定的系统状态比之单变量的情况, 在理论上并无矛盾, 在应用中却相当于增加了限定条件. 状态空间预报的一个关键就在于如何确定系统状态及其邻近态. 所以下面述及的多变量状态空间预报法将取得的应用效果, 是可以预期的.

设有一组变量序列 $X_j(i), i=1, 2, \dots, N$ 表示时间, $j=1, 2, \dots, J$ 表示不同变量. 取滞后参数 τ 和子空间维数 $E_j, j=1, 2, \dots, J$. E_j 不必都相同, 表示不同变量序列的嵌入空间维数可以不同, 它可由应用的预报效果好坏来决定. 多变量状态空间的维数则为 $E = \sum_{j=1}^J E_j$. 类似于单变量的情形, 把多个变量的序列同时嵌入到这个 E 维空间的形式可表示为:

$$P_k = \{X_j(k + k_j \tau)\}, k_j = 0, 1, \dots, E_j - 1, j = 1, 2, \dots, J$$

P_k 可视为 E 维空间一个点. 顺序取 k , 就形成 E 维空间中随时间演变的轨迹.

类似地可拓展前面提到的预报规则. 设预报起点为 P_0 , 选取 P_0 附近一个以 r_j 为半径的 E 维“球”域 S_0 . 其中 r_j 作用于不同变量时可取不同值, 这是因为不同变量的单位和观测精度不同. 设 S_0 内各个邻近 P_0 的点经过 Δt 时间后演变为点集 S'_0 , 则令 S'_0 的中点

P_0 为预报终点.

理论上利用单个时间序列完全可能构造出逼近原物理系统的相空间演变特征,但要求有很长的观测序列,以保障足够高的相点密度.现有大多数气象观测序列不能满足这一要求.应用中采用多个变量序列的优越性在于:

(a) 在多变量状态空间中选取当前状态的历史相似状态的约束条件比单变量情形更严格,从而可排除某些因随机误差造成的虚假的相似状态点,这对提高预报水平显然是有利的.

(b) 所选变量之间总是存在一定的物理或经验上的联系,据此可以建立某些子空间之间的特殊映射关系,如后面将用到的定向预报.这使多变量状态空间预报法具有较大灵活性,便于改善预报模型.

3 应用分析

3.1 单变量情形

预报效果由预报序列和实测序列间的相关系数和符号相关系数(分别记为 C 和 C_s)来衡量.构造预报模型时只用到前述区域序列的前 1000 个值(自 1951 年第一旬始).后 400 多个旬值用于检验预报效果.上述 C 和 C_s 都将由 400 多对预报和观测值得.

显然,嵌入空间维数不同,预报效果也有差别.图 2 表明 $E=3$,即构造三维状态空间可以得到最佳预报效果.此时华东和华北两区的预报相关分别达 0.61 和 0.60,符号相关达 0.73 和 0.64. $E \geq 6$ 预报效果明显变差.理论上用高维空间描写低维系统是不成问题的.但应用中把有限资料序列嵌入高维空间时形成的相点密度相对较低,且高维空间中距离关系更为复杂,确定历史相似状态及其演变轨迹也就更易于受到资料误差的影响.因此图 2 的结果并不能理解为:考虑系统变量个数越多反而使系统的可预报性下降了.系统的可预报性是不会因观测资料处理方法而改变的,只是由于资料数量和质量的

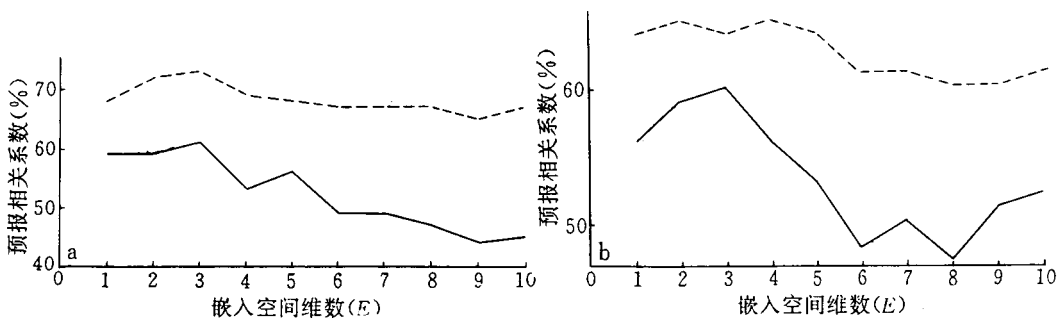
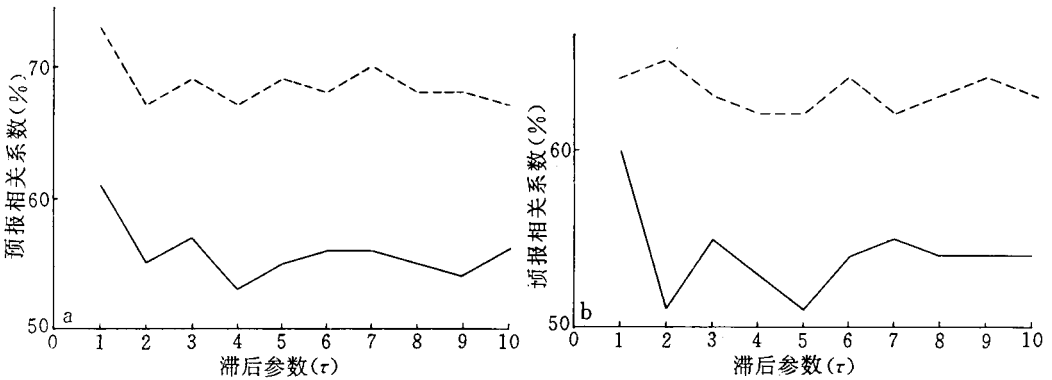


图 2 华东(a)和华北(b)旬雨量变差预报相关(实线)和符号相关(虚线)(滞后参数 $\tau=1$)

Fig. 2 Forecast effects for the dekad-to-dekad rainfall variation in (a) East China (b) North China (solid line: forecast correlation, dashed line: sign-correlation, time-lag parameter $\tau=1$, E : embedded dimensional numbers)

图 3 同图 2 (空间维数 $E=3$)Fig. 3 As in Fig. 2 ($E=3$)

问题,使我们充分利用系统可预报性进行实际预报的能力受到了限制.图 2 的结果或许恰好说明:只要我们恰当地引进更多资料,旬雨量的预报潜力还可进一步地发掘出来.

滞后参数 τ 对预报效果的影响较小.图 3 显示: $\tau=1$ 时效果最好.从现有有关研究来看,多数研究者倾向于取较小的 τ 值,很少有取 $\tau>3$ 者.这和理论上要求 τ 应充分大以确保所重构空间不同轴向相对独立的要求似乎相悖,这也体现了应用研究重实效的原则.

从图 2 和图 3 可以判断:旬雨量变差的演变是遵循一定的动力学规律的.因为一个纯随机系统的状态空间预报应具有 0 相关.而本文的 400 多对值预报相关达到 0.6 以上,决非偶然.图 4 给出华东和华北两区的 $C-\Delta t$ 和 $C_g-\Delta t$ 曲线.图中虚线为一阶自回归

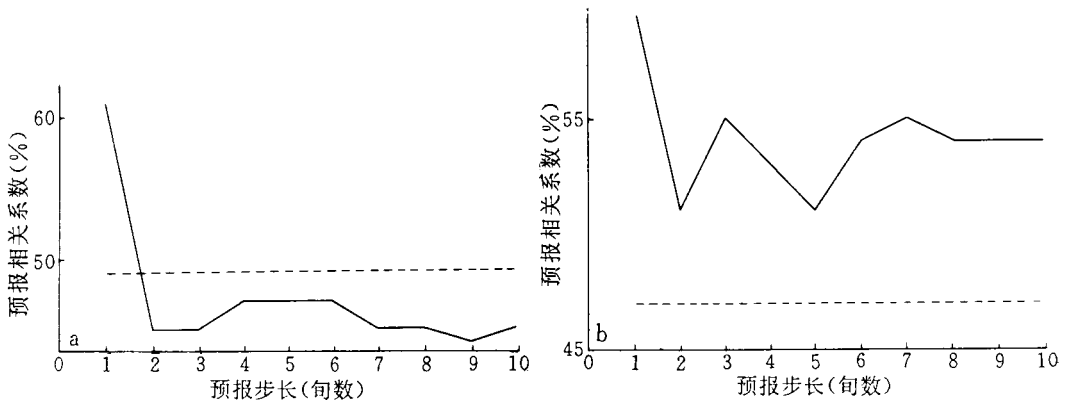


图 4 华东(a)和华北(b)旬雨量变差的预报效果随预报步长的变化($E=3, \tau=1$, 虚线为一阶自回归预报一步的相关水平, 预报值和观测值共 439 对)

Fig. 4 The forecasting correlation of dekad-to-dekad rainfall variation with step ($E=3, \tau=1$ dashed line: 1-step-predicting correlation level of the one-order-autoregressive model) in (a) East China (b) North China (439 couples of data)

方程预报一个时间步长(即一句)的预报相关水平, 回归方程仍由前 1000 个序列点导出. 这样就提供了一个参考, 以比较状态空间预报和传统的概率预报的效果差异.

由图 4 可见, 预报步长 Δt 由 1 变到 2 以后, 预报效果迅速下降. $\Delta t \geq 2$ 以后的预报相关系数大致维持在一个水平上振荡. Sugihara 和 May^[9]认为: 预报相关随预报步长迅速下降说明存在混沌吸引子, 且其可预报时间可由预报效果开始速降的起点 Δt 来判断. 图 4 似表明, 就旬雨量问题而言, 可预报时间为一句, 即恰为观测序列的分辨率. 尽管应用中完全可以构造多旬预报模型, 图 4 却反映了降水预报困难所在的一个理论背景.

3.2 多变量情形

图 5 为两个区域的单变量序列状态空间预报所得到的预报-实测相关图. 图中落在斜率为 1 的直线上的点为实测结果. 预报点则散落在该直线附近, 其偏离直线越远表示预报误差越大. 两个区的共同点是: 对实测较大的绝对值的预报大多数均偏小, 大正值情形尤其严重(大正值表示下旬雨量很大或当旬雨量很小, 大负值表示当旬雨量很大或下旬雨量很小). 图 5 说明: 实况出现大雨时预报水平较差.

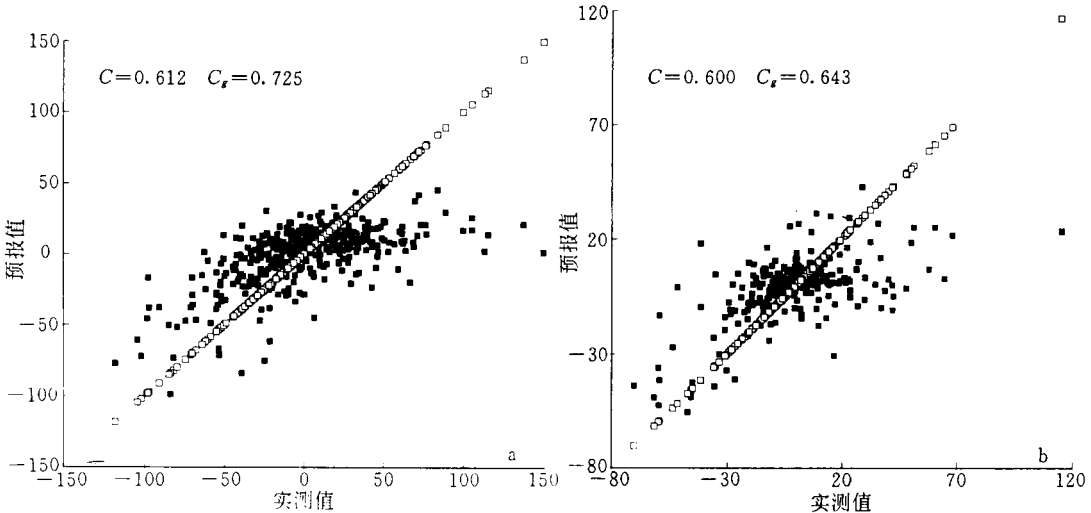


图 5 华东(a)和华北(b)旬雨量变差的单变量状态空间预报相关图(共有 437 对预报-观测值)

Fig. 5 Correlation diagram for single-variable-state-space forecasting of dekad-to-dekad rainfall variation in (a) East China, (b) North China (437 couples of data for observation-prediction)

经验表明: 大雨事件前旬既可能多雨也可能少雨. 这就是说: 从雨量序列本身很难找到大雨事件的前兆信号. 这是单变量状态空间不易报好大雨事件的原因, 而多变量状态空间则提供了一条客观的途径. 作为初步的试验, 本文除雨量资料外, 仅考虑旬气温序列来构造多变量状态空间预报.

本文的多变量状态空间包含上述两变量序列所对应的两个三维子空间. 第三个三维子空间由旬序号(n_i)、雨量变差(X_i)及某个气温距平函数 $f(T_i)$ 构成, 用以设计定向预报, 这是针对大雨事件而作的. 为此, 需要考察大雨事件的一些基本特点.

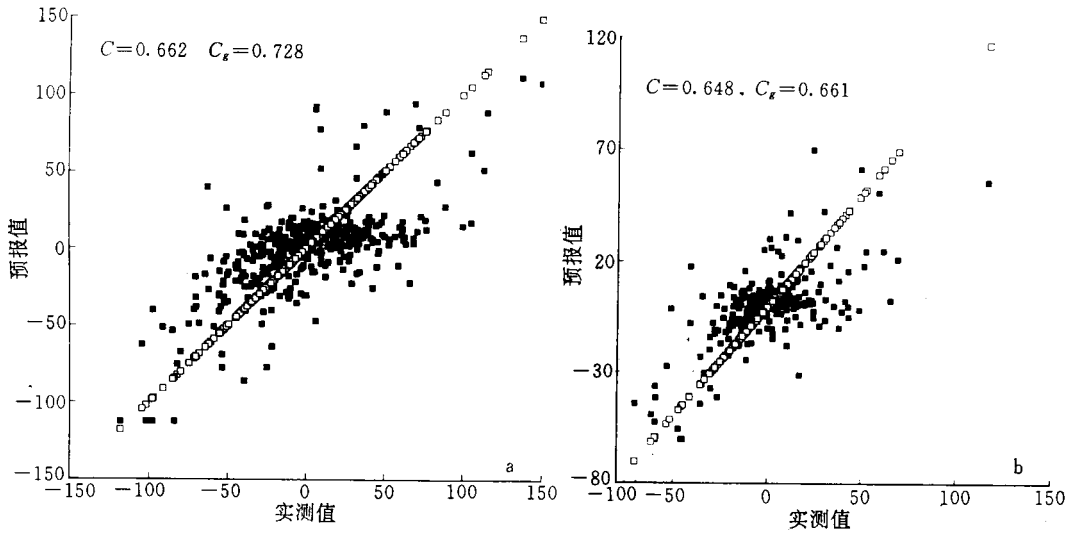


图6 同图5, 但用多变量状态空间

Fig. 6 As in Fig. 5, but for multi-variable-state-space

我们考察了华北雨量变差序列前 1000 个观测中距平量大于 3 个标准差的极端事件, 记为 X_{i+1} 。不难发现, 所有 X_{i+1} 所在旬的序号 n_{i+1} 介于 19 和 24 之间, 即大雨事件仅发生于 7、8 月间。其中不少 X_{i+1} 的前一旬值 X_i 较小而前两旬气温 $T_i + T_{i-1}$ 较高。由此确立如下定向预报关系: 若 $X_i < X_0$, $f(T_i) \equiv T_i + T_{i-1} > T_0$, 且 n_i 位于 $[18, 23]$, 则令 $X_{i+1} = X_m$ 。其中 X_0 、 T_0 和 X_m 为所考察事件有关量的上、下限或平均值。还可以找到其它一些定向预报关系, 应用中可依据其各自对预报效果的提高与否来取舍。从理论上说, 极端事件的发生频率较小, 其在相空间中处于相点密度极低的区域, 在这种区域确定相点轨迹是困难的。在多变量状态空间中利用不同变量之间的物理或经验关系作定向预报, 是解决这一困难的手段之一。而在应用中, 经验性的预报规则是常用的。多变量状态空间法不失为一种对经验规则加以软件化或进一步客观化的工具。

图 6 显示了多变量状态空间预报的相关型式。比较单变量的图 5, 结果已有明显改善。两个区域的预报相关分别提高了约 5%。符号相关的改善程度较小, 说明预报改善主要体现在量值上。这点从图 6 和图 5 中也可明显看出: 前者在大距平值的预报方面大大优于后者。这对于气象预报是特别重要的。应该指出, 图 6 所示的结果中一些较小的值被预报得过大。这说明本文确定的大距平事件的前期特征状态并非都会演变成大距平事件。由于本文仅用雨量和气温两个变量, 其所构成的状态空间在确定预报点之“邻近点”时条件限制仍不够严格。实际上大气降水除了能从其自身及气温历史中找到一定的规律外, 还与其它许多因子如大气环流等有关。因此, 如果能结合气压或风场等变量构造状态空间, 则可望进一步改善预报结果。本文仅提供一个初步的应用范例。

4 小 结

单变量状态空间预报分析表明:对区域旬雨量变差序列而言, Sugihara-May 意义上的可预报时间为一句. 这反映了雨量预报问题困难所在的一个理论背景. 但本文考虑的两个区域 400 多对观测-预报值的相关可达 0.60 以上, 符号相关分别达 73% 和 64%, 这对雨量预报而言又是相当好的结果. 说明雨量虽不象气温、气压等缓变要素那样具有强持续性, 它仍是遵循一定的动力学规律的, 且此规律可在一定程度上为我们的方法所把握.

结合气温要素构造多变量状态空间作预报, 结果得到改善. 由于改善的结果主要是关于那些大距平事件的, 其意义对气候预报是值得强调的. 另一方面, 由于本文仅考虑气温和雨量两组变量构造状态空间, 相对于单变量状态空间情形, 预报效果的改善程度有限. 对雨量问题, 如能结合大气环流等因素构造更多变量状态空间, 则可望进一步改善预报效果.

有关重构系统相空间的理论在应用中要求有很长的资料序列, 目的在于使相空间中相点密度足够高而不致过分歪曲原系统演变轨迹. 但应用中人们所关心的大距平事件(如大旱或大涝)不常有, 在相空间中恰好对应状态点较稀疏的区域, 这就给准确定义其演变轨迹(也即预报)造成了困难. 多变量状态空间法在一定程度上克服了理论与应用之间的这一障碍. 在这个意义上, 本方法对于缺乏长序列资料的气候分析领域具有应用价值. 对于历来强调多要素分析的气象预报, 本方法则提供了一条易于将传统方法客观化和软件化的途径.

致谢: 作者工作期间曾与叶笃正先生进行过详尽讨论, 王晓春在资料方面提供了帮助, 谨一并致谢.

参 考 文 献

- 1 Nicolis C & Nicolis G. Is there a chaotic attractor in climate? *Nature*, 1984, **311**: 529~532.
- 2 Keppen C & Nicolis C. Global and local property of the chaotic attractor in 500 hPa height field over western Europe. *J. Atmos. Sci.*, 1989, **46**: 2356~2370.
- 3 刘式达, 郑祖光, 林振山. 气候层次理论. LASG 论文集之二, 李崇银等主编. 北京: 科学出版社, 1992. 1~5.
- 4 严中伟. 华北旱涝变化的一些混沌性质分析. *气象学报*, 1995, **53**(2): 234~237.
- 5 Yan Z, Ye D & Wang C. Climatic jumps in historical flood/drought chronology in central China. *Climate Dynamics*, 1992, **6**: 156~163.
- 6 Ye D & Yan Z. Climatic jumps in history. *Climate Variability*. Ye D, et al. (eds.). Beijing: China Meteorological Press, 1993. 3~14.
- 7 Yan Z. Evolution of the spectra of historical flood/drought fluctuations. *China Science Bulletin*, 1994, **39**(8): 664~669.
- 8 Cuomo V, Serio C, Crisciani F & Ferrard A. Discriminating randomness from chaos with application to a weather time series. *Tellus*, 1994, **46A**: 299~313.
- 9 Drosowsky W. Analog (nonlinear) forecasts of the Southern Oscillation Index time series. *Weather and Forecasting*, 1994, **9**: 78~84.

- 10 Sugihara G & May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 1990, **344**: 734~741.
- 11 Lin Z, Liu J & He J. The self-organizing methods of long-term forecasting. *Meteor. Atmos. Phys.*, 1994, **53**(3~4): 155~160.
- 12 Van den Dool H M. Searching for analogues, how long must we wait? *Tellus*, 1994, **46A**: 314~324.

MULTI-VARIABLE STATE-SPACE METHOD AND ITS APPLICATION IN DEKAD RAINFALL FORECAST

Yan Zhongwei

(Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029)

Abstract

Combining the two streams of thoughts, i. e. the state-space reconstructed from single variable time series and traditional multi-variable analysis, a multi-variable state-space forecasting method was developed and applied to dekad rainfall prediction for two regions in eastern China. The new method, by considering temperature and seasonality in rainfall-state-space, improved the prediction correlation by 5% (for 437 pairs of data). The improvement remarkably deals with large rainfall deviations, thus having particular significance to meteorological prediction. Great potential of further improvement through combining more reasonable variables in state-space could be expected.

Key words: Multi-variable; State-space; Dekad-to-dekad rainfall variation; Forecasting.