

# 基于 RBF 神经网络水文过程均值遍历特征分析<sup>\*</sup>

王红瑞<sup>1\*\*</sup> 林欣<sup>2</sup> 周冬荔<sup>2</sup> 刘昌明<sup>1</sup> 刘来福<sup>2</sup>

1. 北京师范大学水科学研究院——水沙科学教育部重点实验室, 北京 100875; 2. 北京师范大学数学科学学院, 北京 100875

**摘要** 提出了利用系统聚类、自相关趋势图和径向基函数神经网络对水文过程的均值遍历性特征进行研究, 并以黄河流域的兰州降水站、汉江流域的安康降水站为例, 进行了具体的计算和分析, 结果表明: 兰州站和安康站的 8 月份降水历时过程具有均值遍历性特征的趋势. 最后利用离差系数印证和解释了所得结论的合理性. 同理, 也可以将这一研究思路应用于径流、蒸发等其他水文过程的遍历特征分析.

**关键词** 水文过程 遍历性 平稳性 自相关趋势图 神经网络

## 1 问题的提出

水文现象随时间而变化, 称之为水文过程, 水文过程具有随机性的特点<sup>[1]</sup>. 平稳过程则是指概率特性不随时间而变化的随机过程, 由于水文数据在其获得过程中特有的不确定性, 故所得为随机过程的一个现实. 如果所研究的水文过程是平稳的, 而且能够用一个现实去求统计特征(如均值、方差等), 则会是很有意义的, 并给水文预报等工作带来很大的方便. 由于水文变量随时间变化是极为错综复杂的, 普遍受到随机因素的影响. 要用随机过程理论处理或描述这一系统, 必须要知道它的统计特征, 如相均值函数  $m(t) = E\{\xi(t)\}$ , 相关函数  $R(s, t) = \text{Cov}(\xi(s), \xi(t))$  等. 但在现实中, 通常只是知道过程的一个样本函数, 亦即, 若考虑的是随机序列  $\{\xi(t)\}$ , 只有它的一个有限的时间记录  $\{\xi(t); t=1, 2, \dots, N\}$ . 能否用有限的资料估计总体过程的统计特征? 这在统计学中首先被提出, 并表为相平均(总体均值)与时间平均(样本均值)最终趋于一致. 这种性质称作系统的各态历经性或遍历性(ergodic)<sup>[2,3]</sup>. 由于遍历性的研究不易入手, 在水文研究领域, 鲜有涉及对遍历存在性的探讨. 刘

昌明认为水量交换系统中的水量要素的空间分布与时间分配之间, 存在着遍历性, 即受制于各态历经性规律, 这一规律说明自然界的时间和空间特性上的相似性<sup>[4]</sup>. 近年, 有个别学者用指数权 Markov 链探索梅雨强度的预测问题, 得出梅雨强度状态有遍历性的结果<sup>[5]</sup>. 但是, 这种有限状态相互转化的 Markov 链的遍历性并不同于多数学者研究时间序列中的遍历性. 通常, 时间序列中的遍历性指的是平稳过程遍历性, 解决随机过程在时间上取平均与在空间上取平均是相等的. 目前, 专门探讨水文过程时间序列遍历性的文章尚未推出. 然而, 作为研究许多时序问题的假设基础, 对于遍历性问题本身的研究, 很多学者认为是很有意义, 并且是必不可少的<sup>[4, 6-10]</sup>. 本文在系统聚类的基础上, 利用自相关趋势图和神经网络提出一种具有普适性特点的均值遍历性分析方法.

### 1.1 各态历经性(遍历性)定义

各态历经性(遍历性)理论中说明, 若随机过程(或序列)的均值和协方差函数都有遍历性, 则称此随机过程有遍历性. 但是关于协方差函数的遍历性, 往往涉及过程的 4 阶矩, 一般很难验证, 因此

2007-01-08 收稿, 2007-05-17 收修改稿

\* 国家自然科学基金资助项目(批准号: 50579003)

\*\* E-mail: henryzsr@bnu.edu.cn

©1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

本文只讨论随机过程(或序列)的均值遍历性<sup>[11]</sup>。

安康地区 12 个月份降水量聚类结果:

### 1.2 均值遍历性定义

设 $\{\xi(t); t = 1, 2, \dots\}$  是一随机序列, 考虑样本均值序列 $\{M_T, T = 1, 2, \dots\}$ , 其中  $M_T = \frac{1}{T} \sum_{t=1}^T \xi(t)$ , 如果  $\lim_{T \rightarrow \infty} D(M_T) = 0$ , 则样本均值序列是遍历的, 即过程 $\{\xi(t)\}$  具有均值遍历性。

- 1, 2, 3, 11, 12 月
- 4, 5, 6, 7, 10 月
- 8 月
- 9 月

## 2 降水时间序列的遍历特征分析

选取黄河流域兰州降水站点和汉江流域安康站的降水资料, 两个降水站点降水资料记录十分完备、序列长, 其中, 兰州站(103. 70°, 35. 90°)有历时 51 年(1951 年 1 月—2001 年 12 月)的数据, 安康站(109. 03°, 32. 72°)有 70 年(1929 年 7 月—1998 年 6 月)的降水记录数据, 进行时间序列的均值遍历性质分析。

两个地区 12 个月份均分为 4 类, 由于降水特征不同, 分类结果亦不同。

### 2.1 数据的初步分析

降水量是随降水季节的变化而变化的, 不同降水季节的降水量是有显著差异的。在此视历年相同月份的降水数据来自同一母体。

### 2.3 平稳性与遍历性分析

由于逐月的降水序列可能存在周期性的影响, 而不能满足平稳过程要求其均值函数是常数的要求, 因此将具有相同降水特征的月份作为一个随机过程, 考察其平稳性与遍历性。所以, 上述的聚类结果正好为 8 个需研究的随机过程。

从数据形态分析来看, 兰州站和安康站的降水量随降水季节的变化而呈现出不同的变化, 夏季的降水最多, 春秋的降水则相对较少, 冬季的降水量是最少的。安康站的降水量明显高于兰州站。

#### 2.3.1 分析方法

(1) 分析样本序列的自相关函数, 判断序列是否平稳的。

### 2.2 历时降水数据的聚类分析

视每一个月份作为一个变量, 采用系统聚类的方法对历时 12 个月份的降水量进行聚类, 分别研究具有相同降水特征的月份的遍历性质。由于系统聚类法的聚类原则决定于样品间的距离(或相似系数), 通过  $R^2$  统计量、半偏  $R^2$  统计量、伪  $F$  统计量、伪  $t^2$  统计量来决定类个数, 由此反映每一步聚类的前后类内离差平方和的变化情况, 从而表明聚类效果的优劣。

(2) 考虑样本均值序列 $\{M_T, T = 1, 2, \dots\}$ ,  $M_T = \frac{1}{T} \sum_{t=1}^T \xi(t)$ , 及均值  $M_T$  的方差序列  $D(M_T)$ , 并做出  $M_T, D(M_T)$  的曲线图。

(3) 运用径向基神经网络模型对  $D(M_T)$  函数进行仿真与预测, 确定  $D(M_T)$  随  $T$  的增长逐渐趋于  $\infty$  的变化趋势, 从而检验原序列 $\{\xi(t)\}$  是否具有均值遍历性。

#### 2.3.2 兰州站降水过程平稳性与遍历性分析

##### (1) 自相关与平稳性分析

依据平稳过程的基本定义, 对一过程序列作出是否平稳的判断是十分困难的, 在此利用构成时间序列的每个序列值  $\xi_t, \xi_{t-1}, \dots, \xi_{t-k}$  之间的相关关系形成自相关。

聚类结果如下:

兰州站 12 个月份降水量聚类结果:

- 1, 2, 3, 4, 10, 11, 12 月
- 5, 6, 9 月
- 7 月
- 8 月

$$r_k = \frac{\sum_{t=1}^{n-k} (\xi_t - \bar{\xi})(\xi_{t+k} - \bar{\xi})}{\sum_{t=1}^n (\xi_t - \bar{\xi})^2} \quad (1)$$

序列的平稳性可以用自相关分析图来判断: 如果序列的自相关系数快速的趋于 0, 即落入随机区间, 时序是平稳的, 反之则是非平稳的<sup>[12, 13]</sup>。

图 1 为兰州站四个过程降水序列的自相关函数图。

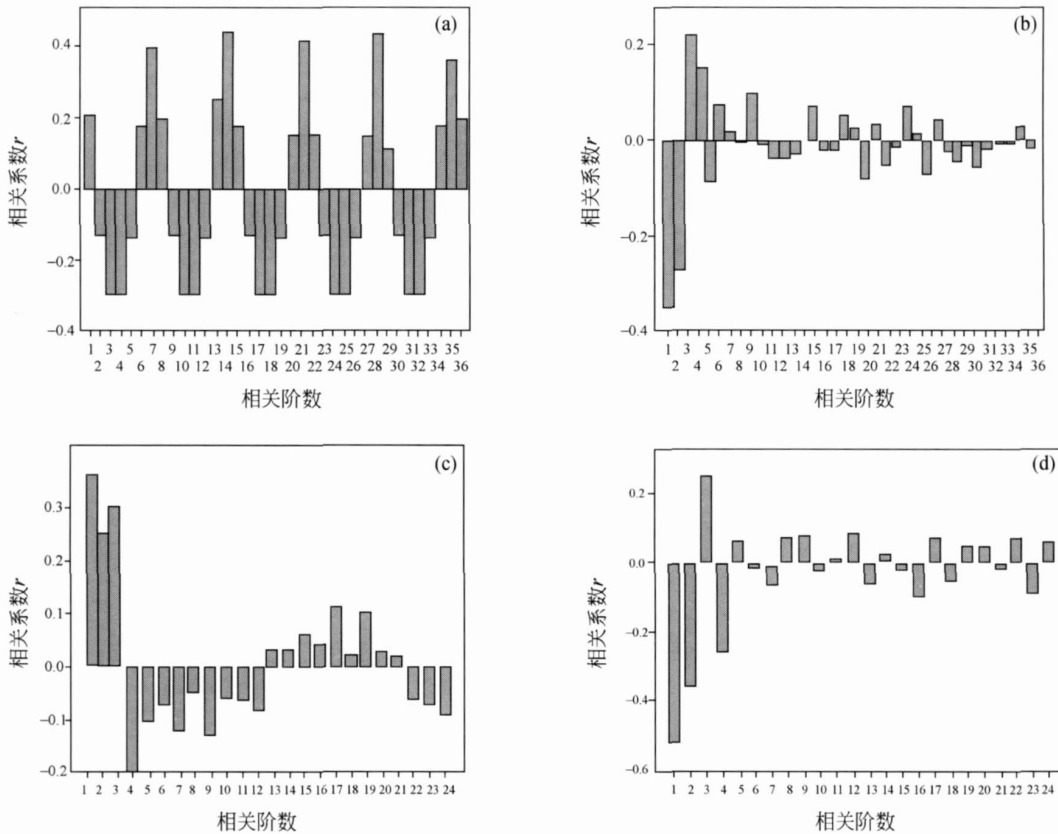


图 1 兰州站各类月份序列自相关图

(a) {兰州站 1, 2, 3, 4, 10, 11, 12 月}; (b) {兰州站 5, 6, 9 月}; (c) {兰州站 7 月}; (d) {兰州站 8 月}

由图 1 可以看出:

{兰州站 1, 2, 3, 4, 10, 11, 12 月} 序列是非平稳的;

{兰州站 5, 6, 9 月} 序列, {兰州站 7 月} 序列, {兰州站 8 月} 序列是平稳的, 即后 3 个时间序列具备继续考察遍历性特征的条件.

(2) 遍历性分析

遍历特征计算公式如下:

$$M_T = \frac{1}{T} \sum_{t=1}^T \xi(t) \quad (2)$$

$$D(M_T) = \frac{\sum_{t=1}^T (M_t - \bar{M}_t)^2}{T} \quad (3)$$

如果  $\lim_{T \rightarrow \infty} D(M_T) = 0$ , 则样本均值序列是遍历的, 过程  $\{\xi(t)\}$  具有均值遍历性.

图 2 为 {兰州站 5, 6, 9 月}, {兰州站 7 月},

{兰州站 8 月} 三个样本序列的  $M_T$ ,  $D(M_T)$  随  $T$  变化的曲线图.

由图 2 可以看出, 由于样本量的有限性, 不能保证  $T \rightarrow \infty$ , 所以  $D(M_T)$  的变化趋势不能够完全确定, 于是需对  $D(M_T)$  拟合并进行预测.

由于  $D(M_T)$  曲线是非线性的, 在此采用径向基函数 RBF 神经网络进行模拟<sup>[14, 15]</sup>. 径向基函数 RBF 神经网络 (即径向基网络, 或者 RBF 网络) 具有较高的运算速度和外推能力, 并且能使网络具有较强的非线性映射功能. 因此, RBF 网络特别适合于非线性较强的序列.

模拟过程如下:

(1) 将原始数据  $x$  做归一化处理<sup>[16, 17]</sup>, 得到新的数据  $x_1$ ,  $x_1 = x / \max(x)$ .

(2) 建立一个 3 层的 RBF 网络<sup>[18]</sup>, 输入层的神经元为  $n_1$  个, 输出层的神经元为  $m$  个;

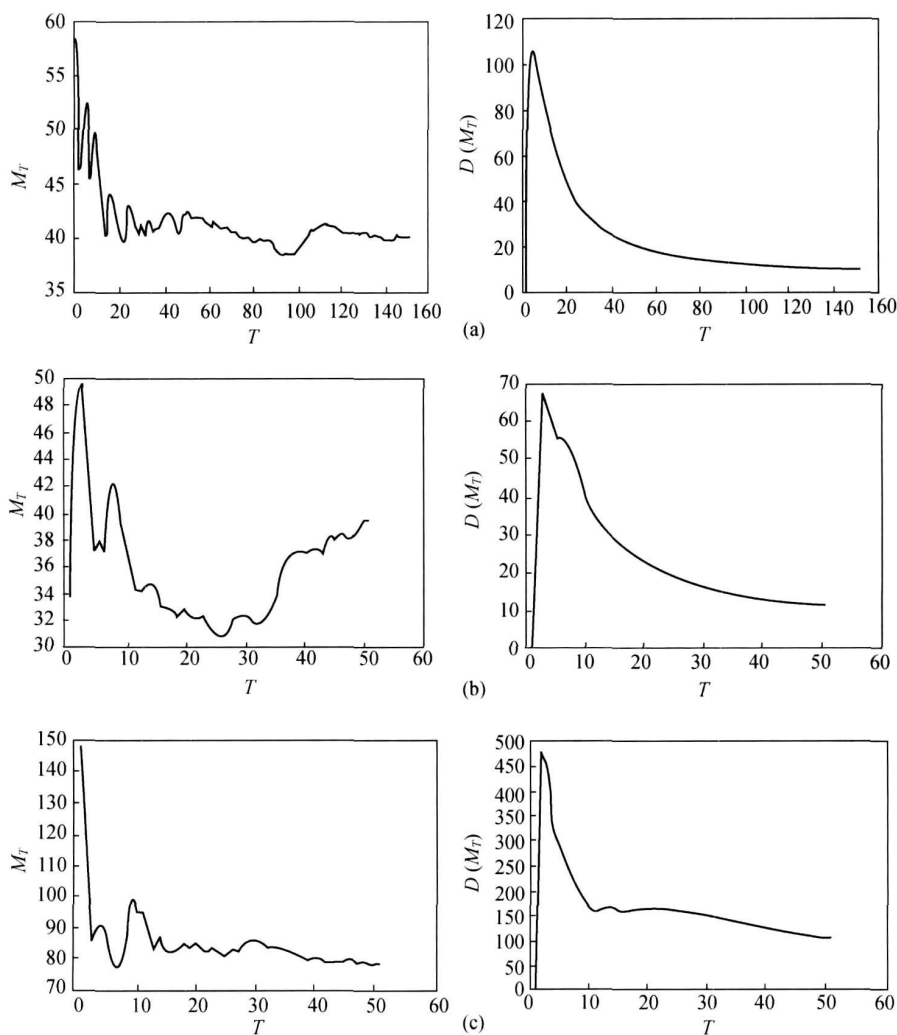


图 2 兰州站各类月份序列  $M_T$ ,  $D(M_T)$  随  $T$  变化的曲线图

(a) {兰州站 5, 6, 9 月}; (b) {兰州站 7 月}; (c) {兰州站 8 月}

(3) 利用 RBF 网络建立前  $n_1$  个数据与第  $n_1+1$  个数据之间的关系, 使这种关系和原始数据具备很高的拟合度;

(4) 连续地作前向预测, 并判断这个序列最终能否达到 0.

模拟及预测结果见图 3.

需特别指出的是, 图 3 中细线为原始序列加进预测序列后的拓展序列, 粗线为预测序列, 并向左平移  $k$  个单位,  $k$  为原始序列长度.

由图 3 可以看出, 只有 {兰州站 8 月} 序列的  $D(M_T)$  曲线函数在  $t=320$  时达到零, 因此得出结论: {兰州站 8 月} 降雨序列具有均值遍历性特征.

2.3.3 安康站降水过程平稳性与遍历性分析 同理, 对安康站的降水进行遍历特征分析.

(1) 自相关与平稳性分析

采用类似于 2.3.2(1) 的分析可知,

{安康站 1, 2, 3, 11, 12 月} 序列是非平稳的;

{安康站 4, 5, 6, 7, 10 月} 序列, {安康站 8 月} 序列, {安康站 9 月} 序列是平稳的, 即具有研究遍历性的前提.

(2) 遍历性分析

安康站三个样本序列的  $M_T$ ,  $D(M_T)$  随  $T$  变化的曲线图如图 4 所示.

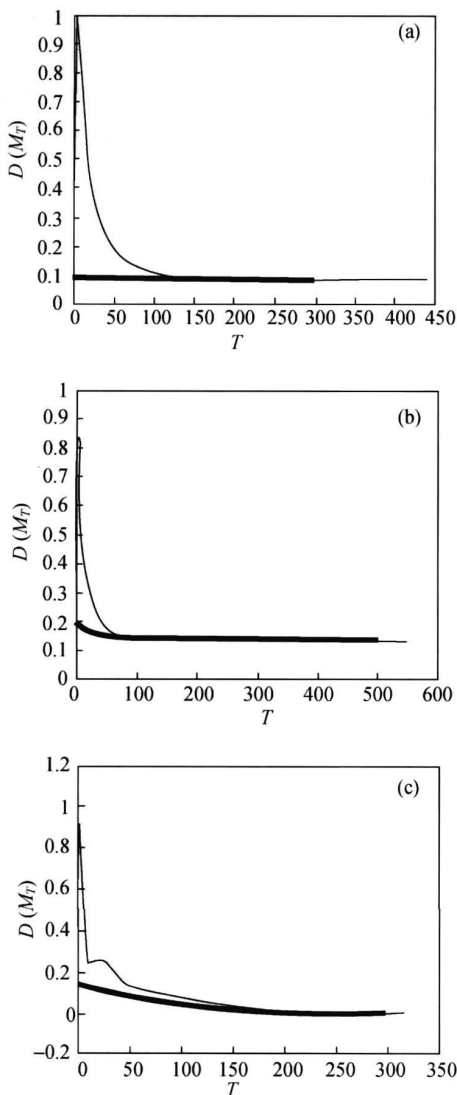


图 3 兰州站各类月份序列  $D(M_T)$  预测曲线  
 (a) {兰州站 5, 6, 9 月}; (b) {兰州站 7 月};  
 (c) {兰州站 8 月}

由图 4 可以看出, 由于样本量的限制, 同样仍不能判断解  $D(M_T)$  在  $T \rightarrow \infty$  时的变化趋势, 所以仍利用径向基函数 RBF 神经网络对  $D(M_T)$  进行仿真模拟并预测, 计算结果如图 5 所示。

由图 5 可以看出, {安康站 8 月} 序列的  $D(M_T)$  曲线函数在  $t = 337$  时达到零, 因此得出结论: {安康站 8 月} 降水序列具有均值遍历性特征。

### 3 印证性分析

由于随机变量分布愈是集中, 其规律性应愈明

显, 反之, 则较差。利用水文过程分析中常用的离差系数<sup>[3,4]</sup> (一种用于衡量随机变量离散程度的指标) 作印证讨论。

以降水量样本的均值表示降水序列随机变量的数学期望, 用其样本标准差表示该随机变量的均方差, 分别计算兰州站与安康站各类月份的变差系数, 结果见表 1。

表 1 兰州站与安康站各类月份降水量离差系数表

兰州站	离差系数	安康站	离差系数
1, 2, 3, 4, 10, 11, 12 月	1.4636	1, 2, 3, 11, 12 月	1.0384
5, 6, 9 月	0.6112	4, 5, 6, 7, 10 月	0.6554
7 月	0.4938	8 月	0.5938
8 月	0.5627	9 月	0.6357

由表 1 可以看出, 两站有平稳性的各类月份, 较无平稳性的月份有更小的离差系数, 其离差系数主要集中在 0.45—0.65 的范围内。而无平稳性各类月份的离差系数值均大于 1.0。由此说明, 离差系数小的各类月份更可能具有平稳性, 即具有更好的规律性, 这与我们预先的猜想不谋而合。然而, 另外值得注意的是: 虽 {兰州站 7 月} 离差系数小于 {兰州站 8 月}, 但 {兰州站 7 月} 降水均值不具有遍历性特征; 而具有遍历性的 {安康站 8 月} 降水均值的离差系数并不显著小于其他平稳月份的离差系数。由此反映, 水文序列离差系数所体现的序列规律性并不能完全涵盖遍历性所反映的规律。因此, 基于离差系数讨论的均值遍历性研究是对水文时序规律更深刻的挖掘。这进一步佐证遍历性研究的深刻性和重要性。

### 4 讨论与结论

对水文过程开展各态历经性(遍历性)分析是一个值得探索的课题, 也是十分困难的, 仅在个别文献上提到水文过程可能存在遍历性特征<sup>[4,19-21]</sup>, 但从未有正式关于遍历性报道的研究成果。

利用兰州、安康两个降水站逐月的历时降水序列作为研究对象, 作者尝试开展了这一方面的研究工作, 由于降水历序列具有有限性、单轨迹和较强的准周期性特点, 无法满足遍历特征的平稳性前提, 因此, 在研究过程中, 本文提出了一些新的思路 (1) 利用系统聚类将每个月份作为一个变量,

根据降水量对其进行聚类, 将具有相同降水特征的月份视作一个整体, 根据时序排定样本, 构成连续时间序列, 讨论它的平稳性和均值遍历性问题, 从水文过程的角度来看, 亦是十分合理的. (2) 提出利用自相关趋势图确定降水时序的平稳性, 避开了

利用平稳过程基本定义检验时间序列平稳性的困难. (3) 提出利用径向基函数(RBF)神经网络, 来模拟并预测样本均值的方差序列, 克服了降水序列有限性的缺陷. (4) 利用各类月份的离差系数作印证性分析.

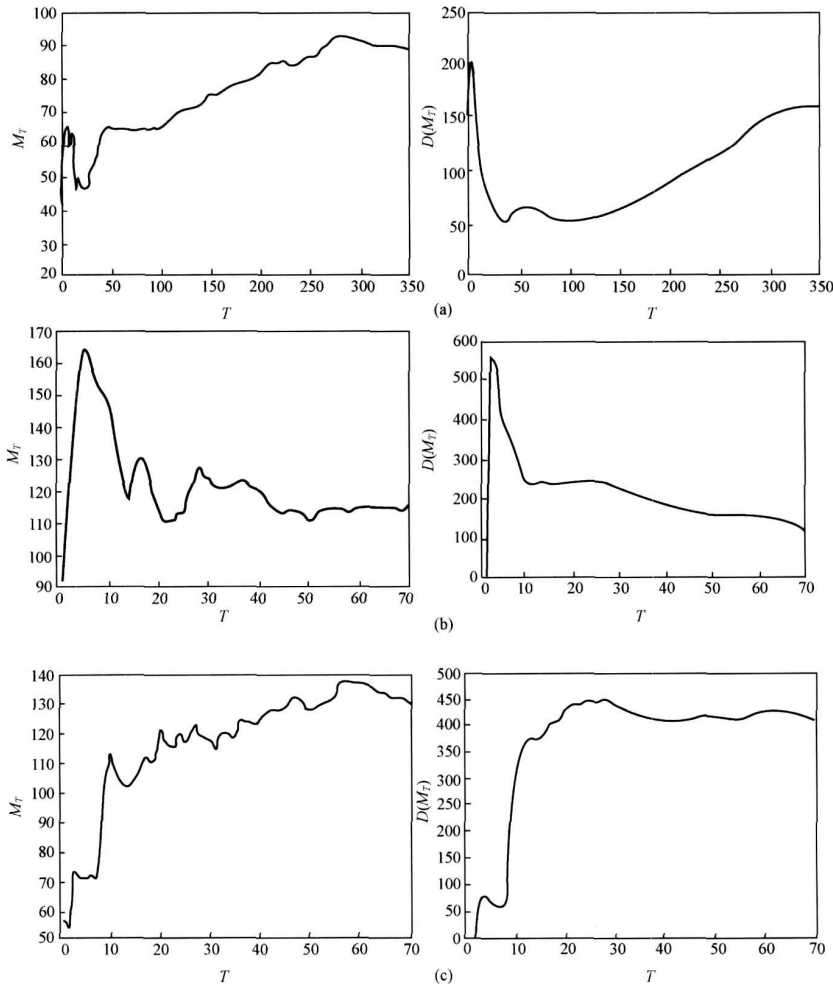


图 4 安康站各类月份序列  $M_T$ 、 $D(M_T)$  随  $T$  变化的曲线图

(a) {安康站 4, 5, 6, 7, 10 月}; (b) {安康站 8 月}; (c) {安康站 9 月}

另外, 需要说明的是径向基函数(RBF)神经网络是一种性能良好的前向神经网络模型, 只要中心选取得当, 仅需要很少的神经元就能获得很好的逼近效果, 而且计算量小, 学习速度快<sup>[22]</sup>. 若遇到更少数据量的情况, 研究者更应慎重选取 RBF 神经网络的中心, 求得更好的逼近效果. 由于具有遍历性特征的数据不太可能出现强烈的波动<sup>[23]</sup>, 因此, 用

RBF 神经网络来预测遍历性的后续变化, 是很好的选择.

从水文学与统计学的角度来说, 兰州站和安康站历年 8 月份的降水过程的均值具有很强的规律性和稳定性.

利用相同的思路与模型, 也可以对径流, 蒸发等其他水文过程的均值平稳性、遍历性特征进行研

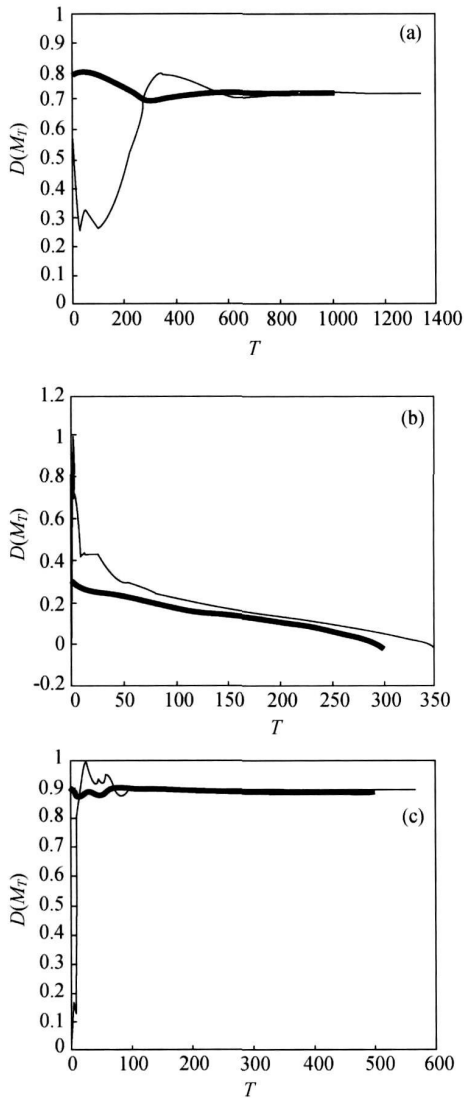


图 5 安康站各类月份图序列  $D(M_T)$  预测曲线

(a) {安康站 4, 5, 6, 7, 10 月}; (b) {安康站 8 月}; (c) {安康站 9 月}

究。如果能对某一水文过程的自协方差进行遍历性分析, 从而探讨该水文过程的遍历性特征, 将是更有意义但同时也是难度更大的工作。

参 考 文 献

- 1 Wang HR, Ye LT, Liu CM, et al. Problems in wavelet analysis of hydrologic series and some suggestions on improvement. Progress In Natural Science, 2007, 17(1): 80—86
- 2 Chick S, Shortle J, Gelder PV, et al. A model for the frequency of extreme river levels based on river dynamics. Structural Safety, 1996, 18(4): 261—276
- 3 华东水利学院. 水文学的概率统计基础. 北京: 水利出版社,

- 1981, 170—173
- 4 刘昌明. 水量转换的若干问题. 见: 刘昌明, 任鸿遵. 水量转换实验与计算分析. 北京: 科学出版社, 1988, 3—21
- 5 夏乐天. 梅雨强度的指数权马尔可夫链预测. 水利学报, 2005, 36(8): 988—993
- 6 Aldo F, Igor J. Can we determine the transverse macrodispersivity by using the method of moments? Advances in Water Resources, 2005, 28(6): 589—599
- 7 Hsu KC. The influence of the log-conductivity autocovariance structure on macrodispersion coefficients. Journal of Contaminant Hydrology, 2003, 65(1-2): 65—77
- 8 李 森, 陈家军, 叶慧海, 等. 随机方法在地下水水分及溶质运移中的研究进展. 水资源研究, 2005, 26(4): 22—25
- 9 Mitosek HT. On stochastic properties of daily river flow processes. Journal of Hydrology, 2000, 228(3-4): 188—205
- 10 丁 晶, 邓育仁. 随机水文学. 成都: 成都科技大学出版社, 1988, 12, 30—32
- 11 方兆本, 缪柏其. 随机过程. 合肥: 中国科学技术大学出版社, 2002, 82—91
- 12 Cline DBH, Huay-min Pu. Verifying irreducibility and continuity of a nonlinear time series. Statistics & Probability Letters, 1998, 40: 139—148
- 13 Cline DBH, Huay-min Pu. Stability of nonlinear AR(1) time series with delay. Stochastic Processes and their Applications, 1999, 82: 307—333
- 14 李月英, 中东日, 陈义俊, 等. 基于 RBF 神经网络的非线性系统的预测. 计算机测量与控制, 2006, 14(3): 319—321
- 15 Murat A, Cigizoglu HK. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. Environmental Modelling & Software, 2007, 22(1): 2—13
- 16 祖家奎, 曾庆丰. 归一化径向基函数网络的结构优化策略. 计算机仿真, 2002, 19(3): 43—45, 56
- 17 Jang JSR, Sun CT, Mizutaru E. Neuro-Fuzzy and Soft Computing. Prentice Hall, 1997, 21—24
- 18 Norgaard M. Neural network based system identification toolbox. 00-E-891, Department of Automation, Technical University of Denmark, 2000
- 19 Duan JQ, Goldys B. Ergodicity of stochastically forced large scale geophysical flows. <http://ijmms.hindawi.com>. [2006-06-16]
- 20 Demetris K. Stochastic Simulation of Hydrosystems. <http://www.itia.ntua.gr/getfile/541/2/2004EncycStochSimulPP.pdf>. [2006-06-16]
- 21 番成正. 优势水流对坡地安定影响之研究——以草蛉为例. 博士论文, 国立成功大学, 资源工程学系, 1993
- 22 韩 敏, 王 晨, 席剑辉. 基于改进 RBF 神经网络的非线性时间序列预测. 仪表仪器学报, 2003, 24(4): 574—575, 581
- 23 周佩玲, 陶小丽, 傅忠谦, 等. 改进 RBF 神经网络用于降雨量预测. 小型微型计算机系统, 2001, 22(2): 244—246