

WD-RBF Model and its Application of Hydrologic Time Series Prediction

Dengfeng Liu, Dong Wang*, Yuankun Wang

Key Laboratory of Surficial Geochemistry, Ministry of Education; Department of Hydro-sciences, School of Earth Sciences and Engineering, Nanjing University
Nanjing, 210046, China

Lachun Wang, Xinqing Zou

School of Geographic and Oceanographic Sciences, Nanjing University
Nanjing, 210046, China

* Corresponding Author E-mail: wangdong@nju.edu.cn

Received 15 September 2013

Accepted 12 December 2013

Abstract

Accurate prediction for hydrological time series is the precondition of water hazards prevention. A method of radial basis function network based on wavelet de-noising (WD-RBF) was proposed according to the nonlinear problem and noise in hydrologic time series. Wavelet coefficients of each scale were calculated through wavelet transform; soft-threshold was used to eliminate error in series. Reconstructed series were predicted by RBF network. The simulation and prediction of WD-RBF model were compared with ARIMA and RBF network to show that wavelet de-noising can identify and eliminate random errors in series effectively; RBF network can mine the nonlinear relationship in hydrologic time series. Examples show that WD-RBF model has superiority in accuracy compared with ARIMA and RBF network.

Keywords: Hydrologic time series, RBF network, Wavelet de-noising, Water hazards

WD-RBF 模型在水文时间序列模拟预测上的应用

刘登峰¹, 王栋^{*1}, 王远坤¹, 王腊春², 邹欣庆²

1. 表生地球化学教育部重点实验室, 南京大学地球科学与工程学院水科学系, 南京, 210046

2. 南京大学地理与海洋科学学院, 南京 210046

*通讯作者 Email: wangdong@nju.edu.cn

摘要: 对水文时间序列科学的模拟预测是有效防范水灾害的前提。针对水文时间序列的非线性和噪声污染问题, 提出了一种基于小波消噪理论的径向基函数网络(WD-RBF)方法。通过小波变换得到水文时间序列各尺度下的小波系数, 利用软阈值消噪技术消除序列随机误差。重构小波还原水文时间序列, 将消噪后的序列进行的 RBF 网络模拟预测。将 WD-RBF 模型与 ARIMA 模型和 RBF 网络进行对比分析, 小波消噪能够对序列中的随机误差进行有效识别并消除, RBF 网络能够挖掘水文时间序列中的非线性映射关系。实例表明 WD-RBF 方法在序列模拟预测上表现出优于 ARIMA 模型和 RBF 的性能, 具有更高的准确性。

关键词: 水文时间序列, RBF 神经网络, 小波消噪, 水灾害

1. 引言

在全球气候极端变化的背景下, 以旱涝为代表的水灾害的频繁发生严重威胁着人类的生命财产安全。水文时间序列作为水文系统的数据化体现, 对其加以准确的模拟预测是有效防范水灾害, 保证流

域和地区可持续发展的重要前提。由于影响水文系统的因素错综复杂, 难以用精确的数学语言来描述, 导致水文时间序列预测建模异常困难。

针对水文时间序列的非线性特征, 人工智能领域内的神经网络方法能够较好地予以解决。径向基函数神经网络(RBF-NN)因具有较强的非线性拟合能

力和局部逼近的特性,近年来被应用于时间序列的模拟预测中。李军和刘君华(2005)采用基于卡尔曼滤波算法的新型广义 RBF 网络预测模型;张军峰和胡寿松(2007)提出一种基于高斯基的距离度量,并联合了输入输出聚类的策略;柳海涛、孙双科、刘之平等(2005)在激发函数中引入 Sign-d 函数,以改善模型的稳定性和泛化能力;杨晓华、杨志峰、沈珍瑶等(2005)用自相关技术分析时间序列的延迟特性,用以确定 RBF 网络的输入、输出向量;刘俊萍和畅明琦(2007)采用最近邻聚类学习算法确定径向基函数的宽度、选取聚类中心和权值;路剑飞和陈子燊(2010)利用引入滞后因子的 DE-RBF 方法,探讨建立多步预测模型的可行性,等等。

除了非线性特征以外,水文时间序列另一个无法回避的问题为潜在的噪声污染。水文系统受自然及人为因素干扰,实测序列往往受到不同程度的噪声污染,对模拟预测工作产生了不利影响。对于水文时间序列存在的噪声污染问题,传统的维纳滤波、卡尔曼滤波等只适用于线型系统,且严格依赖于状态空间函数的建立;而 Fourier 变换仅适用于平稳时间序列(王文圣,丁晶,李跃清,2005)。小波消噪(WA)是消除非平稳时间序列噪音的有效工具。近年来许多学者对小波消噪方法进行了研究和改进:Coifman 和 Wickerhauser (1992) 基于 Shannon 熵提出了小波函数选择的依据和方法;Lou 和 Hu (2003) 基于二进制离散小波变换和 Karhunen-Loeve 变换提出了用于分析非平稳宽带信号的消噪方法;桑燕芳和王栋(2008)将小波变换中常用的 7 个小波系共 54 个小波函数作为选择范围,探讨建立合理选择小波函数的依据和方法,等等。

针对水文时间序列的非线性和噪声污染问题,提出了 WD-RBF 模拟预测方法。引入小波函数对水文序列进行小波变换,并对各层小波系数进行阈值消噪。消噪处理后的序列代入 RBF 网络进行建模预测。以黄河花园口站和利津站的年径流序列,北京市和南京市的年降水序列为例,对 WD-RBF 模型与 ARIMA 模型、RBF 网络进行对比分析,验证方法的准确性。

2. 基本理论

2.1 RBF 网络

人工神经网络包括单层前馈网络,多层前馈网络和递归网络(S.Haykin, 2004)。径向基函数(RBF)网络是一种前馈式网络,具有唯一最佳逼近点,保证全局收敛,因而具有较高的训练速度和仿真效率。

RBF 神经网络基于径向对称的基函数,如式(1)所示高斯核函数,实现输入空间到隐含层的非线性变换。通过线性变换将隐含层数据映射至输出空间,如式(2)所示。

$$\phi(\mathbf{x}, \mathbf{c}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

式(1)中, \mathbf{c}_i 为神经元的核函数中心向量; σ_i 为核函数的宽度参数。

$$\mathbf{Y} = \mathbf{W}\Phi \quad (2)$$

式(2)中, \mathbf{Y} 为 $r \times 1$ 输出向量, \mathbf{W} 为 $r \times n$ 阶权重矩阵, Φ 为 $n \times 1$ 向量。

2.2 小波阈值消噪

小波消噪技术始于小波阈值消噪方法的提出(Donoho, 1995)。此后阈值消噪的方法广泛应用于光谱分析(高国荣,刘艳萍,潘琼,2012),流量预测(窦慧丽,刘好德,吴志周等,2009),信号处理(赵莉,冯稷,翟光杰等,2005)等领域。

阈值消噪的思想即对小波分解后的各层系数中模大于和小于某阈值的系数分别处理,然后对系数进行反变换,重构出消噪后的信号。阈值确定准则有:

(1) 固定阈值

选择式为:

$$T = \sigma_n \sqrt{2 \ln N} \quad (3)$$

其中 σ_n 为噪声强度, N 为小波系列的长度或尺寸。

(2) Stein 的无偏风险阈值

将某一层小波系数的平方由小到大排列,得到一个向量 $W = [w_1, w_2, \dots, w_n]$, ($w_1 \leq w_2 \leq \dots \leq w_n$)。由此计算风险向量 $R = [r_1, r_2, \dots, r_n]$ 。其中,

$$r_i = \frac{n - 2i + (n - i)w_i + \sum_{k=1}^i w_k}{n} \quad (4)$$

以 R 中最小元素 r_b 作为风险值,由 r_b 的下标找到对应的 w_b , 则阈值:

$$T = \sigma \sqrt{w_b} \quad (5)$$

式中 σ 为噪声强度。

(3) Heuristic SURE 阈值

比较两个变量:

$$\eta = \sigma \left[\sum_{j=1}^n |x_j|^2 - n \right] / n \quad (6)$$

$$\text{crit} = \sigma \sqrt{\frac{1}{n} \left(\frac{\ln n}{\ln 2} \right)^3} \quad (7)$$

如果 $\eta < \text{crit}$, 则选用固定阈值; 否则取固定阈值和无偏风险阈值中的较小者作为本准则的阈值。

固定阈值处理小波系数的缺点是会歪曲原始信号包含的弱特征成分, 从而引起重构信号的失真; 而当序列信噪比较小时, SURE 估计会有较大误差。Heuristic SURE 阈值是固定阈值和 SURE 阈值的综合, 所选择的是最优预测变量阈值。本文选择 Heuristic SURE 作为阈值确定准则。

3. WD-RBF 水文时间序列模拟预测方法

3.1 基本思路

(1) 小波消噪阶段:

①小波分解。选择合适的小波函数, 确定小波分解的层次 N , 对水文时间序列进行 N 层小波分解。序列自身的变化特性是小波函数选择的重要影响因素。而对于小波层数的确定, 若层数偏小, 低频子序列仍是多个时间尺度上信号的混叠, 且由于序列能量压缩不明显, 无法有效去除噪声; 若层数过大, 较大分解层上会产生一些缺乏物理意义的虚假信号, 且会造成误差累积等问题 (Chou C M, 2011)。基于序列变化特点选择截频性能出色小波函数 *dmey*(桑燕芳, 王栋, 2008), 并依据自相关白化检验确定分解层数 (李炜, 陈晓辉, 毛海杰, 2009)。②阈值消噪。对 $1 \sim N$ 层的每一层高频系数选择合适的阈值 T 进行量化处理, 得到去噪后的序列成分。因为在硬阈值方法中, 信号在 T 处是不连续的, 会给重构序列带来一定振荡。采取 Heuristic SURE 的软阈值方法对实例序列进行阈值处理。③小波重构。根据小波分解的第 N 层的低频系数和经过处理后的第 1 层到第 N 层的高频系数进行重构, 得到消噪后的水文时间序列。

(2) RBF 网络模拟预测阶段:

①确定输入层神经元数。对原始数据进行归一化处理。分别计算实例序列的自相关系数, 判断序列自身的延迟特性。取自相关系数显著趋于 0 所对应的延迟阶数 K 作为 RBF 网络输入层神经元数。②训练网络。在 RBF 网络训练中, 隐层神经元数的确定是一个关键问题, 过去的做法是使其与输入矢量数相等, 显然过多的隐层单元数增加了运算的负

担。提出改进方法: 从 1 个神经元开始训练, 通过检查输出误差使网络自动增加神经元; 每循环计算一次后, 用使网络产生最大误差所对应的训练样本作为权值向量产生一个新的隐层神经元, 然后重新计算, 并检查新网络的误差, 重复此过程直到达到误差要求或最大隐层神经元数为止, 网络训练完成。③模拟预测。将消噪后的实例序列代入 RBF 网络, 以预测步长 K 进行预测模拟, 以检验预测方法的有效性和准确性。

WD-RBF 水文时间序列模拟预测方法流程如图 1 所示。

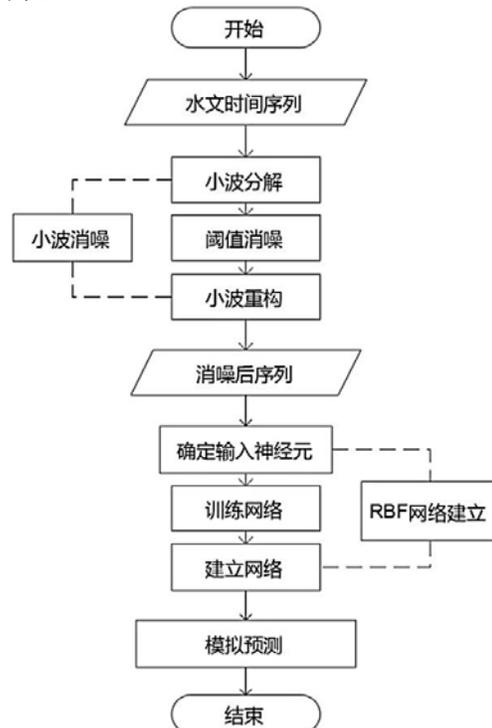


图1 WD-RBF模拟预测方法流程图

3.2 方法实现

选取黄河花园口站 1950-2003 年径流量, 利津站 1950-2003 年径流量, 北京市 1951-2011 年降水量, 南京市 1951-2011 年降水量为实例。序列最后 5 年值作为验证数据不代入建模。

对实测水文时间序列进行 *dmey* 离散小波分解, 得到各个尺度下的小波系数。对各个尺度下的小波系数进行 Heuristic SURE 软阈值消噪处理, 消噪残差序列如图 2 所示。

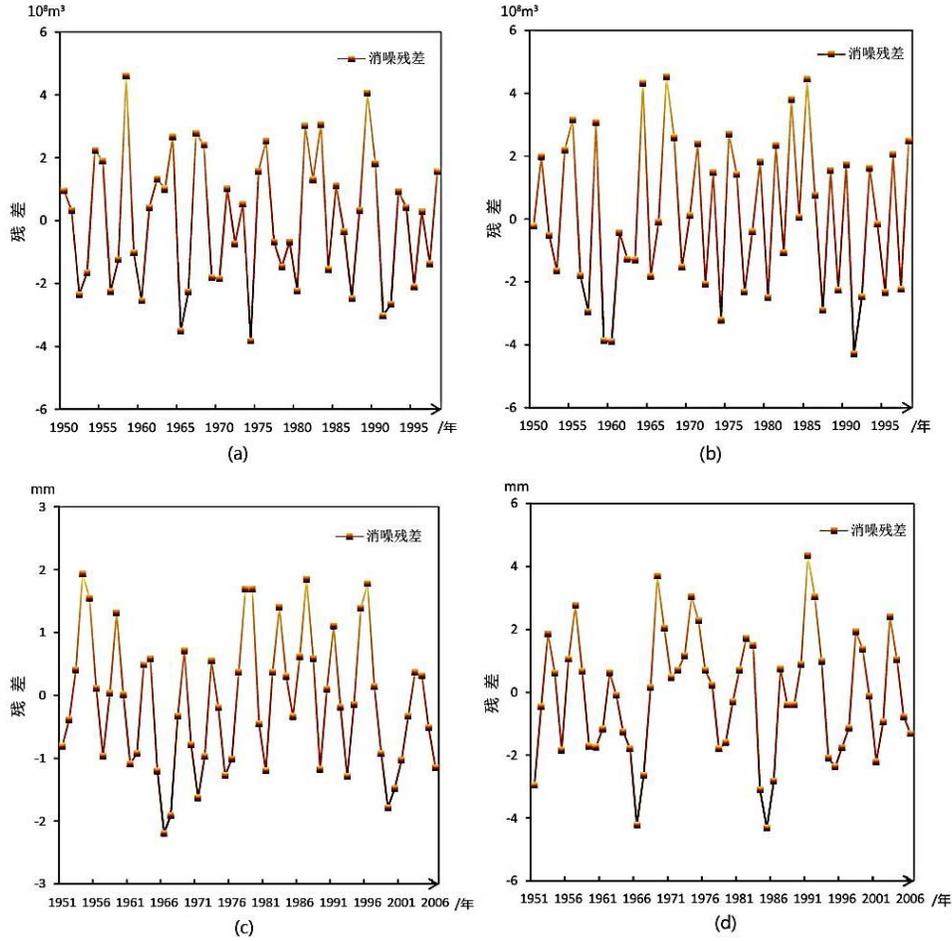


图2 各序列消噪残差图
 (a) 花园口站年径流 (b) 利津站年径流 (c) 北京年降水 (d) 南京年降水

图 2 显示各消噪残差呈现出一个矩形散点图且没有明显的模式,表明没有出现明显偏离噪声随机性外观的情况 (Jonathan D.Cryer, Kung-Sik Chan,2008)。对各消噪之后的序列计算不同延迟阶数 K 下的自相关系数(ACF)。北京市 K -ACF 计算结果如表 1 所示。

表1 北京市年降水量K-ACF表

K	1	2	3	4
ACF	0.1670	0.0895	0.1811	0.1624
K	5	6	7	8
ACF	0.1821	-0.0506	-0.0476	0.0393

表 1 所示, 延迟阶数 K 增大至 6 时, 序列的自相关系数显著降低。因此选取北京年降水量序列的 RBF 网络预测步长 (即输入层神经元个数) 为 5 (路剑飞, 陈子燊, 2010)。同理依次确定各时间序列的网络输入层神经元数, 结果如表 2 所示。

表2 RBF网络输入层神经元数对应表

序列	花园口年径流	利津年径流	北京年降水	南京年降水
神经元数	3	5	5	5

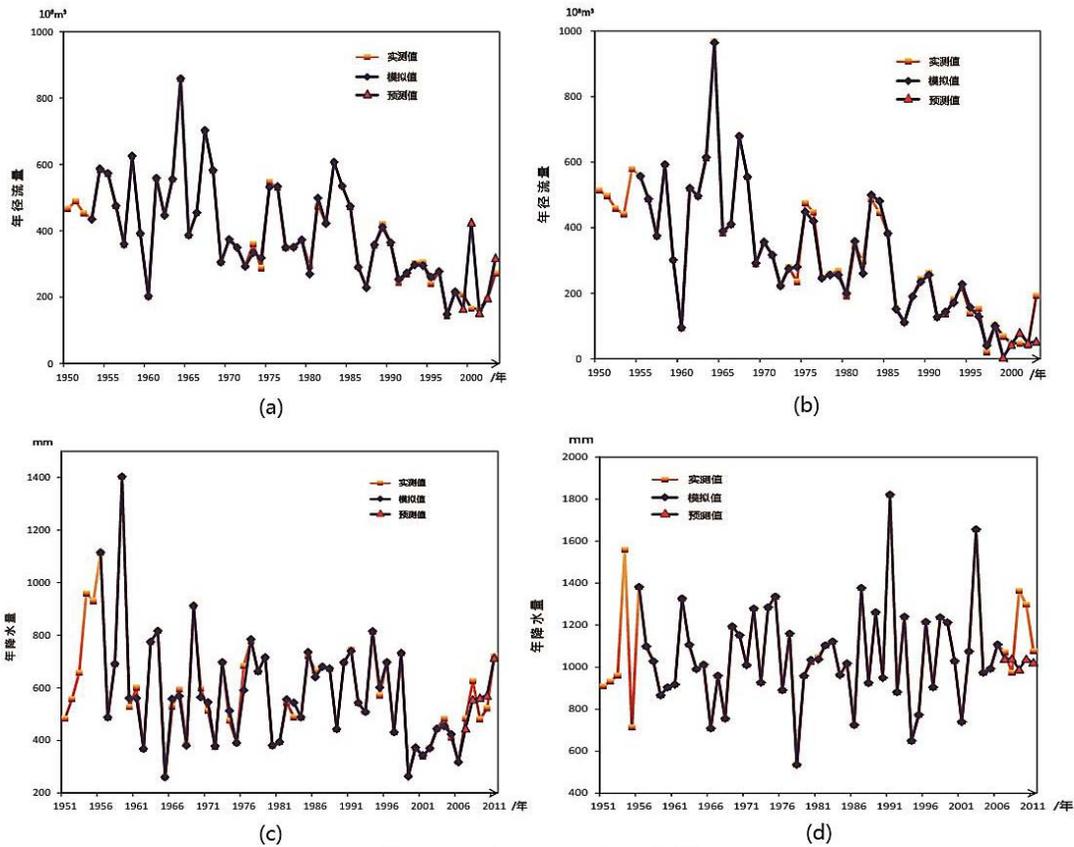


图3 各序列WD-RBF模拟预测图
 (a) 花园口站年径流 (b) 利津站年径流 (c) 北京年降水 (d) 南京年降水

确定输入层神经元个数，训练并建立相应 RBF 网络，运用建立好的 RBF 网络对序列进行模拟预测。各水文时间序列 WD-RBF 模拟预测结果如图 3 所示。

4. 对比分析

选取的对比方法为基于序列线性非平稳假设的

ARIMA 模型及非线性假设的 RBF 网络。对原实际序列直接进行 RBF 网络模拟预测，并建立 ARIMA 模型模拟预测(Box G E P, Jenkins G M, Reinsel G C. 2011)。将 WD-RBF 模型与 ARIMA 模型、RBF 网络的模拟预测结果进行对比。

选取预测的绝对误差 AE(表 3)，预测的相对误差 RE(表 4)和模拟预测的均方误差 MSE(表 5)三组指标对各方法的准确性进行比较分析。

表3 ARIMA/RBF/WD-RBF时间序列预测AE表

预测年份	花园口站年径流序列/ 10^8m^3			利津站年径流序列/ 10^8m^3		
	ARIMA 模型	RBF 网络	WD-RBF	ARIMA 模型	RBF 网络	WD-RBF
1	62.27	47.66	-43.09	100.25	8.61	-63.42
2	53.66	280.63	260.7	36.26	30.99	-4.79
3	79.87	60.17	-12.14	203.54	105.18	34.05
4	107.65	89.58	2.02	69.75	92.63	6.15
5	-67.17	117.02	45.74	36.79	-63.15	-139.13

表4 ARIMA/RBF/WD-RBF时间序列预测RE表

预测年份	北京市年降水序列			南京市年降水序列		
	ARIMA 模型	RBF 网络	WD-RBF	ARIMA 模型	RBF 网络	WD-RBF
1	-19.26%	-8.00%	-7.78%	-25.08%	0.03%	-2.86%
2	-15.25%	-11.35%	-11.13%	-8.85%	9.88%	6.77%
3	-6.53%	16.82%	17.04%	-5.19%	-27.80%	-27.40%
4	20.51%	9.66%	9.42%	-51.58%	-17.63%	-19.93%
5	-20.44%	-1.07%	-0.78%	-29.23%	-1.35%	-4.98%

表5 ARIMA/RBF/WD-RBF时间序列模拟预测MSE表

方法	花园口站年径流序列 /10 ⁸ m ³		利津站年径流序列 /10 ⁸ m ³		北京市年降水序列 /mm		南京市年降水序列 /mm	
	MSE ^①	MSE ^②	MSE ^①	MSE ^②	MSE ^①	MSE ^②	MSE ^①	MSE ^②
ARIMA	41.21	76.47	43.31	108.64	57.57	101.96	343.59	355.58
RBF	1.77	65.23	2.38	31.41	2.93	25.06	0.56	90.70
WD-RBF	1.36	53.69	2.11	31.37	2.94	24.85	0.33	92.73

MSE^①: 模拟值均方误差 MSE^②: 预测值均方误差

由表 5, WD-RBF 模型和 RBF 网络对于水文时间序列的模拟精度较 ARIMA 模型有显著提高。以花园口年径流量为例, ARIMA 模型模拟的均方误差为 41.21, RBF 网络模拟的均方误差为 1.77, 而 WD-RBF 模拟之后的均方误差最小, 为 1.36。从模拟的对比分析可以得出, WD-RBF 模型和 RBF 网络的模拟精度高于 ARIMA 模型, WD-RBF 模型的结果较 RBF 网络有一定程度提高。

对于水文时间序列的预测, 以花园口年径流量为例, 由表 3, ARIMA 模型 5 年预测最小绝对误差为 53.66, RBF 网络的最小绝对误差为 47.66, WD-RBF 模型的最小误差为 2.02。又以北京年降水为例, 由表 4, ARIMA 模型预测相对误差范围在 15%-20% 之间, 最小相对误差绝对值 6.53%, 最大为 20.51%; RBF 网络预测相对误差范围在 10%-15%, 最小相对误差绝对值 1.07%, 最大为 16.82%; 而 WD-RBF 模型的相对误差绝对值为 0.78%, 最大为 17.04%。在序列的预测上, WD-RBF 模型和 RBF 网络的预测精度高于 ARIMA 模型, WD-RBF 模型在 RBF 网络基础上有所提高, 但提高幅度不大。

5. 结论

对水文时间序列加以科学的模拟预测是水灾防范的前提和依据。针对水文时间序列模拟预测中的非线性特征和噪声污染问题, 提出了 WD-RBF 模拟预测模型。RBF 网络能够模拟出水文时间序列中

内在的非线性机制, 小波消噪能够一定程度上识别消除序列中的随机噪声, 达到还原真实序列的目的。实例表明, WD-RBF 模型在模拟预测中的精度显著高于 ARIMA 模型, 并在 RBF 网络基础上有所提高, 因而表现出优于传统 ARIMA 模型和 RBF 网络的性能, 具有更高的准确性。

然而 WD-RBF 网络较 RBF 网络并没有表现出明显的优越性, 其原因是否否与消噪过程有关需要进一步分析论证。针对不同信噪比的水文时间序列, 如何选择合适的小波函数和消噪阈值, 仍值得进一步研究。

致谢

资助项目: 国家自然科学基金 (No. 41071018)、国家重点基础研究发展计划 (2013CB956503)、教育部新世纪优秀人才支持计划 (NCET-12-0262)、教育部博士点基金 (20120091110026, 20100091120059)、江苏省教育厅青蓝工程。

参考文献

1. Powell, M. J. D. Radial basis functions for multi-variable interpolation: A review, IMA Conference on Algorithms for the Approximation of Functions and Data, RMCS Shrivensham, UK, 1985.

2. D. S. Broomhead and D. Lowe. Multivariate functional interpolation and adaptive networks [J], *Complex Systems*, vol. 2, 1988, pp. 321-355.
3. S. Haykin. *Neural Networks: A Comprehensive Foundation*[M]. Macmillan College Publishing Company: Englewood Cliffs, 1994.
4. Donoho, D.L. De-Noising by Soft-thresholding [J]. *IEEE Transactions on Information Theory*, 1995. 41(3): p. 613-627.
5. D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage [J]. *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200-1224, 1995.
6. Box G E P, Jenkins G M, Reinsel G C. *Time series analysis: forecasting and control* [M]. Wiley, 2011.
7. Coifman R R, Wickerhauser M V. Entropy-based algorithms for best basis selection [J]. *Information Theory, IEEE Transactions on*, 1992, 38(2): 713-718.
8. Lou H W, Hu G R. An approach based on simplified KLT and wavelet transform for enhancing speech degraded by non-stationary wideband noise [J]. *Journal of sound and vibration*, 2003, 268(4): 717-729.
9. Chou C M. Wavelet-based multi-scale entropy analysis of complex rainfall time series[J]. *Entropy*, 2011, 13(1): 241-253.
10. Cryer J D, Chan K S. *Time series analysis with applications in R*[M]. Springer, 2008.
11. Zhang Junfeng, Hu Shousong. Chaotic time series prediction based on RBF neural networks with a new clustering algorithm [J]. *Acta Phys. Sin.*, 2007(2):713-719.
张军峰, 胡寿松. 基于一种新型聚类算法的 RBF 神经网络混沌时间序列预测[J]. *物理学报*, 2007(2): 713-719.
12. Li Jun, Liu Junhua. On the prediction of chaotic time series using a new generalized radial basis function neural networks [J]. *Acta Phys. Sin.*, 2005, 54(10): 4569-4577.
李军, 刘君华. 一种新型广义 RBF 神经网络在混沌时间序列预测中的研究[J]. *物理学报*, 2005, 54(10): 4569-4577.
13. Sang Yanfang, Wang Dong. Wavelets selection method in hydrologic series wavelet analysis [J]. *Journal of Hydraulic Engineering*, 2008, 39(3): 295-300.
桑燕芳, 王栋. 水文序列小波分析中小波函数选择方法[J]. *水利学报*, 2008,39(3): 295-300,306.
14. Lu Janfei, Chen Zichen. Salinity Prediction at Modaomen Waterway in Estuary of Pearl River. *Journal of China Hydrology*, 2010(5): 69-74.
路剑飞, 陈子燊. 珠江口磨刀门水道盐度多步预测研究. *水文*, 2010(5): 69-74.
15. Liu Junping, Chang Mingqi. Water Demand Prediction Based on Radial Basis Function Neural Network [J]. *Journal of China Hydrology*, 2007(5): 12-15, 11.
刘俊萍, 畅明琦. 径向基函数神经网络需水预测研究 [J]. *水文*, 2007(5): 12-15, 11.
16. Yang Xiaohua, Yang Zhifeng, Shen Zhenyao, et al. High precision radial basis function network model for predicting water environment nonlinear time series[J]. *Advances in Water Science*, 2005(6): 788-791.
杨晓华, 杨志峰, 沈珍瑶等. 水环境非线性时序预测的高精度 RBF 网络模型[J]. *水科学进展*, 2005(6): 788-791.
17. Liu Haitao, Sun Shuangke, Liu Zhiping, et al. Atomization prediction based on artificial neural networks for flood releasing of high dams [J]. *Journal of Hydraulic Engineering*, 2005(10): 1241-1245. 017.
柳海涛, 孙双科, 刘之平等. 泄洪雾化预测的人工神经网络方法探讨[J]. *水利学报*, 2005(10): 1241-1245.
18. Gao Guorong, Liu Yanping, Pan Qiong. A differentiable thresholding function and an adaptive threshold selection technique for pulsar signal denoising [J]. *Acta Phys. Sin.*, 2012(13): 549-553.
高国荣, 刘艳萍, 潘琼. 基于小波域可导阈值函数与自适应阈值的脉冲星信号消噪[J]. *物理学报*, 2012(13): 549-553.
19. Zhao Li, Feng Ji, Zhai Guangjie, et al. Wavelet transformation for magnetocardiography signal [J]. *Acta Phys. Sin.*, 2005(4): 1943-1949.
赵莉, 冯稷, 翟光杰等. 小波变换在心磁信号处理中的应用[J]. *物理学报*, 2005(4): 1943-1949.
20. Dou Huili, Liu Haode, Wu Zhizhou, et al. Study of Traffic Flow Prediction Based on Wavelet Analysis and Autoregressive Integrated Moving Average Model [J]. *Journal of Tongji University(Natural Science)*, 2009(4): 486-489, 494.
窦慧丽, 刘好德, 吴志周等. 基于小波分析和 ARIMA 模型的交通流预测方法[J]. *同济大学学报(自然科学版)*, 2009(4): 486-489, 494.
21. Li Wei, Chen Xiaohui, Mao Haijie. A Study of Adaptive Optimal Decomposition Level in Threshold De-noising Algorithm Based on Wavelet Transform [J]. *Computer Simulation*, 2009, 26(3): 311-314.
李炜, 陈晓辉, 毛海杰. 小波阈值消噪算法中自适应确定分解层数研究[J]. *计算机仿真*, 2009, 26(3): 311-314.