

基于 WA-SVM 的水库溶解氧预测

罗华军^{1a,2}, 黄应平^{1a}, 刘德富^{1b}

(1 三峡大学 a 化学与生命科学学院, b 土木水电学院, 湖北 宜昌 443002; 2 武汉大学 水利水电学院, 湖北 武汉 430072)

【摘要】 【目的】提出一种基于小波分析-支持向量机(WA-SVM)的水库溶解氧预测模型,以期提高水库溶解氧的预测精度。【方法】通过小波分解,将原始复杂的溶解氧浓度序列分解到不同的高频和低频层次,对每层得到的分解重构序列分别采用支持向量机回归方法进行预测后,合成原始序列的预测值,将该模型应用于桥水库溶解氧浓度序列的预测中,并与单独支持向量机(SVM)回归方法预测结果进行比较。【结果】WA-SVM 方法预测精度较 SVM 方法有较大提高,其平均绝对百分比误差和均方根误差分别为 0.049 37 和 0.345 3,而 SVM 方法的分别为 0.084 93 和 0.631 9。【结论】WA-SVM 方法综合运用了小波分析的多分辨特性和支持向量机的非线性回归功能,能够较准确地预测水库溶解氧浓度。

【关键词】 小波分析;支持向量机;溶解氧;预测模型

【中图分类号】 X832.03;TP181

【文献标识码】 A

【文章编号】 1671-9387(2009)03-0181-06

Predicting on reservoir dissolved oxygen based on WA-SVM

LUO Hua-jun^{1a,2}, HUANG Ying-ping^{1a}, LIU De-fu^{1b}

(1a College of Chemistry & Life Science, b College of Hydroelectric & Civil Engineering, Three Gorges University, Yichang, Hubei 443002, China; 2 College of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, Hubei 430072, China)

Abstract: 【Objective】 To enhance the precision of reservoir dissolved oxygen forecast, wavelet analysis and support vector machine (WA-SVM) forecast model was established. 【Method】 The original complex dissolved oxygen concentration series were decomposed to different layers through wavelet analysis, then each layer forecasted by means of SVM, and finally the forecasting results of the original time series obtained by composition. The model was applied to predict the dissolved oxygen series of Yuqiao reservoir, and compared with the results of SVM. 【Result】 The results indicated that the forecast accuracy of WA-SVM was higher than that of SVM. The mean absolute percentage error (MAPE) and the root of mean square error (RMSE) of WA-SVM method was 0.049 37 and 0.345 3 respectively, whereas MAPE and RMSE of SVM method were 0.084 93 and 0.631 9. 【Conclusion】 WA-SVM method can accurately predict reservoir dissolved oxygen concentration because of comprehensively using multi-distinguishable character of wavelet analysis and nonlinear regression function of support vector machine.

Key words: wavelet analysis; support vector machine; dissolved oxygen; predicting model

当今水库水质污染问题越来越严重^[1-2],而溶解氧(DO)浓度是水体受污染程度和生态环境好坏的重要指标之一^[3]。因此,对水库溶解氧的分析预测就显得十分重要。当前,关于溶解氧的预测方法主

要有多元回归分析法^[4-5]、人工神经网络法^[6-7]、动力学模型法^[3,8]等。由于水库溶解氧变化规律及特征的复杂性和不确定性,上述常规方法不仅需要确定较多的影响因素,而且预测的准确性也不太高。目

* [收稿日期] 2008-04-25

[基金项目] 国家自然科学基金项目(50679038)

[作者简介] 罗华军(1974—),男,湖北武汉市人,副教授,在读博士,主要从事水环境科学数值模拟研究。

E-mail:luohuajun@21cn.com

[通信作者] 刘德富(1962—),男,湖北枝江市人,教授,博士生导师,主要从事生态水工研究。E-mail:dfliu@ctgu.edu.cn

前小波分析(Wavelet Analysis,简称 WA)和支持向量机(Support Vector Machine,简称 SVM)等新方法在水科学领域的应用研究均已取得一定成果^[9-12],将两者结合应用于水质预测,不失为一条新的途径。

20 世纪 80 年代初兴起的小波分析(WA),是一种窗口大小固定但形状可变、时间窗和频率窗均可改变的时频局部化分析方法,具有对信号自适应性和多分辨的特性^[13]。Vapnik^[14]和 Suykens 等^[15]在 20 世纪 90 年代提出的 SVM 方法,是基于统计学习理论的机器学习工具,遵循结构风险最小化原理,能够在对小样本学习的基础上,对其他样本进行快速、准确的拟合预测。

本文将 WA 和 SVM 2 种方法相结合进行水库溶解氧预测,先用 WA 将原始复杂的数据序列分解为几层简单序列的叠加,然后对每层分别采用 SVM 进行预测,最后合成预报结果,并将该方法应用于桥水库溶解氧浓度序列的预测中,同时与单独 SVM 预测结果进行了比较,以期为提高水库溶解氧预测精度提供参考依据。

1 小波分析(WA)

WA 是一种窗口大小固定但形状可变(时宽或频宽可变)的时频局部化分析方法,其关键在于引入了满足一定条件的基本小波函数 $\psi(x)$,用以代替 Fourier 变换中的基函数 $e^{-i\omega x}$ 。

设函数 $\psi(x)$ 为一平方可积函数,即 $\psi(x) \in L^2(R)$,若 $\psi(x)$ 的 Fourier 变换满足允许性条件

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty, \tag{1}$$

则称 $\psi(x)$ 为一个基本小波或小波母函数^[13]。式

(1)中: ω 表示频率, $\hat{\psi}(\omega)$ 为 $\psi(x)$ 的 Fourier 变换。

$\psi(x)$ 经伸缩和平移得到一族函数:

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right), a, b \in R, a \neq 0. \tag{2}$$

式中: $\psi_{a,b}(x)$ 称为分析小波或连续小波, a 为尺度(伸缩)因子, b 为时间(平移)因子。

小波变换分为连续和离散 2 种,实际应用中一般采用离散小波变换(DWT)。将式(2)中的参数 a, b 都取离散值,固定伸缩步长 $a_0 > 1$,平移步长 $b_0 \neq 0$,取 $a = a_0^{-m}, b = nb_0 a_0^{-m}$,从而将连续小波变成离散小波:

$$\psi_{m,n}(x) = a_0^{m/2} \psi(a_0^m x - nb_0), m, n \in Z. \tag{3}$$

对于任意函数 $f(t)$,经小波分析可以将其分解

成高频率序列 d_1 和低频率序列 a_1 ,然后将低频率序列 a_1 进一步分解,如此重复可以得到任意尺度上的高频序列和低频序列(图 1)^[16]。

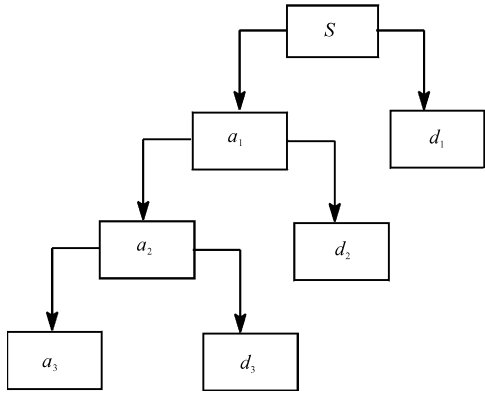


图 1 小波分解示意图

Fig. 1 Schematic diagram of wavelet decomposition

2 支持向量机(SVM)算法

SVM 是近年来出现的基于统计学习理论的、解决多维函数预测的机器学习工具。其遵循结构风险最小化原理,克服了神经网络的过学习问题,在小样本情况下具有良好的外推能力。其主要分析思想是:通过事先选择的非线性映射,将输入向量映射到高维特征空间,在这个空间中构造最优决策函数。在构造最优决策函数时,本算法利用了结构风险最小化原则,同时还巧妙地利用原空间的核函数取代了高维特征空间的点积运算。支持向量机常用的学习算法有 Chunking 算法、分解算法、序列最小优化 SMO 法、最近点快速迭代(NPA)算法和最小二乘法等。

Suykens 等^[17]提出的最小二乘支持向量机(LS-SVM)算法,主要是优化函数与其他方法不同,只有等式约束,而无经典的 SVM 的不等式约束。其原理与算法如下:设样本为 n 维向量,某区域的 l 个样本及其值表示为: $(x_1, y_1), \dots, (x_l, y_l) \in R^n \times R$ 。首先用一个非线性映射 $\varphi(\cdot)$,将样本从原空间 R 映射到特征空间 $\varphi(x) = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_l))$ 。在这个高维特征空间中构造最优决策函数 $y(x) = w \cdot \varphi(x) + b$,这样,非线性估计函数就转化为高维特征空间的线性估计函数。利用结构风险最小化原则寻找 w, b ,即最小化 $R = \|w\|^2/2 + c\xi$,其中: $\|w\|^2$ 为控制模型的复杂度, c 为惩罚因子, ξ 为误差控制函数,即 ϵ 不敏感损失函数。常用的损失函数有线性 ϵ 损失函数、二次 ϵ 损失函数、Huber 损失函数。选取不同的损失函数,可构造不同形式

的支持向量机。最小二乘支持向量机在优化目标中的损失函数为误差 ξ 的二次项,故优化问题为:

$$\begin{cases} \min J(w, \xi) = \frac{1}{2}w \cdot w + c \sum_{i=1}^l \xi_i^2, \\ s. t: y_i = w \cdot \varphi(x_i) + b + \xi_i (i=1, \dots, l). \end{cases} \quad (4)$$

用拉格朗日法求解该优化问题:

$$L(w, b, \xi, a, y) = \frac{1}{2}w \cdot w + c \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l a_i (w \cdot \varphi(x_i) + b + \xi_i - y_i). \quad (5)$$
$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & k(x_1, x_1) + 1/c & \cdots & k(x_1, x_l) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & k(x_l, x_1) & \cdots & k(x_l, x_l) + 1/c \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_l \end{bmatrix}.$$

最后得到非线性模型:

$$f(x) = \sum_{i=1}^l a_i k(x, x_i) + b. \quad (8)$$

式(8)中不同的内积函数目前主要有 3 类:多项式核函数、径向基核函数和 sigmoid 核函数。研究表明,径向基核函数可取得良好的计算效果^[18-20]。本研究取径向基核函数

$$k(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right). \quad (9)$$

3 WA-SVM 预测模型的建立

本文将小波分析和支持向量机算法相结合,建立 WA-SVM 预测模型。首先通过小波分析将原始复杂的水库溶解氧浓度序列,分解为低频序列(趋势项)和高频序列(随机项),根据序列本身的特征,再将低频序列继续分解。对分解得到的各序列先进行归一化处理,然后确定支持向量机的输入向量数和选择支持向量机参数,分别建立最小二乘支持向量机回归模型进行模拟预测。最后再将各分频序列得到的预测值反归一化处理后进行合成重构,即得到整个水库溶解氧的预测值。

4 应用实例

4.1 研究区域概况

于桥水库位于燕山山脉南麓,处于天津市蓟县城东 4 km,距离天津市区 115 km,是引滦入津工程的重要调蓄水库,也是天津市人民生活饮用水和工农业生产用水的水源地。近 10 多年来,于桥水库流域社会经济发展迅速,入库污染负荷急剧增加,目前水库已处于重富营养状态。2000-01-01~2002-07-20,于桥水库坝前溶解氧浓度序列(每月 1 日、10

式中: a_i 是拉格朗日乘子, $i=1, \dots, l$ 。据优化条件,对 w, b, ξ, a 求偏导数,并令其为 0,得:

$$\begin{aligned} w &= \sum_{i=1}^l a_i \varphi(x_i), \sum_{i=1}^l a_i = 0, a_i = c \xi_i, \\ w \cdot \varphi(x_i) + b + \xi_i - y_i &= 0. \end{aligned} \quad (6)$$

定义核函数 $k(x_i, x_j) = \varphi(x_i) \varphi(x_j)$,其中 $k(x_i, x_j)$ 是满足 Mercer 条件的对称函数。根据式(6),优化问题可转化为求解线性方程:

$$\begin{bmatrix} 1 & k(x_1, x_1) & \cdots & k(x_1, x_l) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & k(x_l, x_1) & \cdots & k(x_l, x_l) + 1/c \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_l \end{bmatrix}. \quad (7)$$

日、20 日各监测 1 次)^[21] 如图 2 所示。

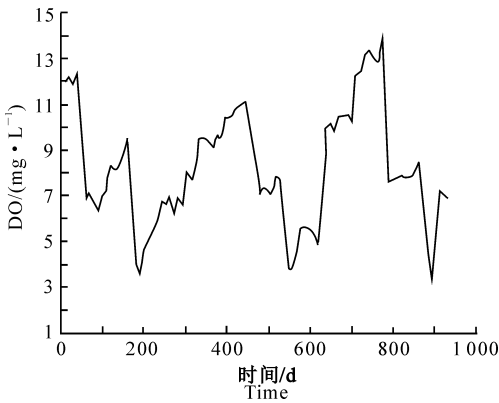


图 2 于桥水库坝前溶解氧浓度序列
Fig. 2 Dissolved oxygen concentration series of Yuqiao reservoir in front of dam

4.2 预测步骤

首先采用二进制 Daubechies 小波系中的‘db4’小波进行多尺度分解,将原始序列 s 分解到第 3 层,得到低频序列 a_3 和高频序列 d_3, d_2, d_1 (如图 3 所示);然后对分解得到的各个序列运用支持向量机回归方法进行预测,得到各自的预测值 a_3' 和 d_3', d_2', d_1' ,采用 MATLAB7.1 软件编程,具体步骤如下:

(1)对数据进行归一化处理,避免各个因子之间的量级差异。

(2)确定支持向量机的输入向量数(序列的历史步数 p),对样本数据 $\{x_i\} = \{x_1, x_2, \dots, x_N\}$ 进行相空间重构。即用前 p 个时刻的浓度值 x_i, x_{i+1}, x_{i+p-1} 预测 $i+p$ 时刻的浓度值 x_{i+p} 。经分析试算,取历史步数 p 为 7。将最后 9 组数据作为测试样本,其余 77 组数据作为学习样本,进行滚动一步预测。

(3)选择支持向量机参数。惩罚因子 c 与径向

基核函数参数 σ 对支持向量机的预测能力有较大影响。采用网格交叉验证法,优选参数为 $c=97.80$, $\sigma=9.962$ 。

(4)利用优化的参数 c 和 σ 建立最小二乘支持

向量机模型,对各序列进行预测。对各序列预测值反归一化处理后,用公式(10)进行合成重构,得预测结果(果如图 4 所示):

$$s'=a_3'+d_3'+d_2'+d_1'。$$
 (10)

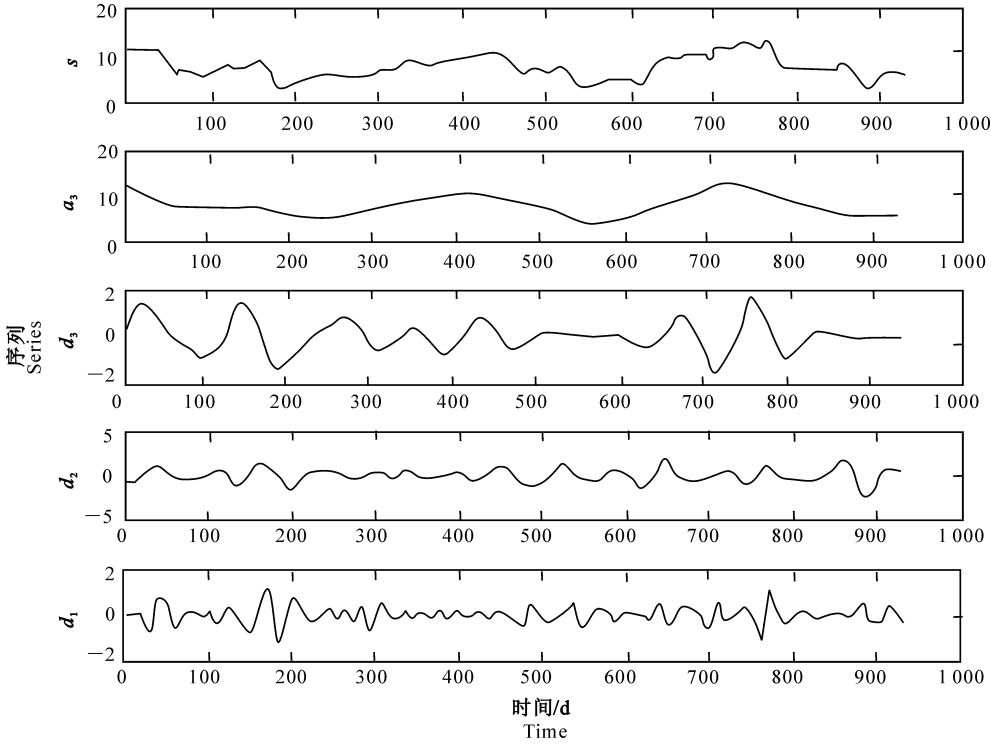


图 3 于桥水库溶解氧浓度序列小波分解过程

Fig. 3 Wavelet decomposition process of dissolved oxygen concentration series in Yuqiao reservoir

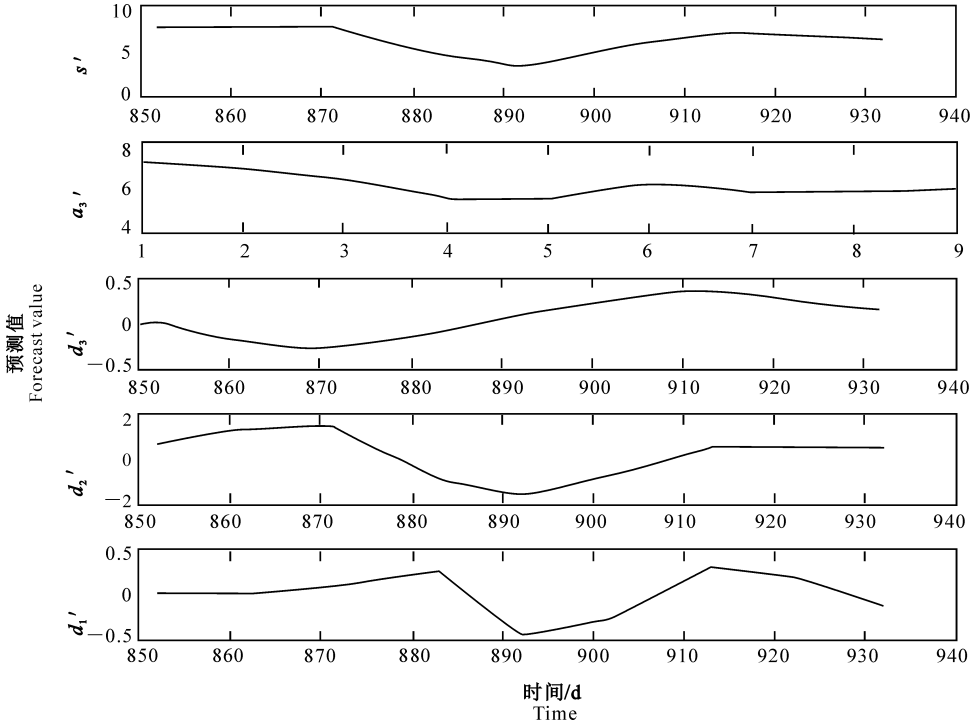


图 4 于桥水库溶解氧各分频序列及合成的预测值

Fig. 4 Forecast value of dissolved oxygen concentration separate frequency series and composition in Yuqiao reservoir

4.3 WA-SVM 与 SVM 预测结果的比较

为了将 WA-SVM 与 SVM 的预测结果进行比较,特对原始序列直接采用支持向量机方法预测,学

习样本、测试样本同上,仍然选择径向基核函数,SVM 参数为 $c=97.80, \sigma=9.962$ 。2 种方法的预测结果与实测值的比较见表 1。

表 1 基于 SVM 和 WA-SVM 的预测结果与实测值比较

Table 1 Contrast between forecast results and survey values based on SVM and WA-SVM

日期 Date	实测值/ (mg · L ⁻¹) Survey value	SVM		WA-SVM	
		预测值/ (mg · L ⁻¹)	相对误差/%	预测值/ (mg · L ⁻¹)	相对误差/%
		Forecast result	Relative error	Forecast result	Relative error
2002-05-01	8.100 0	7.813 5	−3.54	7.730 3	−4.56
2002-05-10	8.509 7	7.546 9	−11.31	7.963 3	−6.42
2002-05-20	7.267 8	7.898 4	8.68	7.715 3	6.16
2002-06-01	4.500 0	5.117 0	13.71	4.755 2	5.67
2002-06-10	3.334 0	3.462 1	3.84	3.472 4	4.15
2002-06-20	5.857 5	4.847 7	−17.24	5.514 8	−5.85
2002-07-01	7.300 0	6.514 0	−10.77	7.074 3	−3.09
2002-07-10	7.061 5	7.418 8	5.06	6.837 4	−3.17
2002-07-20	6.841 3	6.684 7	−2.29	6.474 3	−5.36

采用平均绝对百分比误差 MAPE(式 11)和均方根误差 RMSE(式 12)对 2 种方法的预测效果进行评价^[22]：

$$MAPE=\frac{1}{n}\sum_{i=1}^n\left|\frac{y_i-\hat{y}_i}{y_i}\right|$$
 (11)

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2}$$
 (12)

式中: y_i 为实测值, $\hat{y}_i$ 为预测值, n 为测试样本的数量。

由表 1 可知,WA-SVM 方法的平均绝对百分比误差和均方根误差分别为 0.0493 7 和 0.345 3,SVM 方法的分别为 0.084 93 和 0.631 9。可见,WA-SVM 方法比单独 SVM 方法的预测效果更好,误差更小。

5 结 论

本文运用小波分析的分频功能和多分辨特性,采用遵循结构化风险最小化原则的支持向量机回归预测技术,将两者结合应用到于桥水库溶解氧浓度时间序列的预测中,结果表明,WA-SVM 方法的预测精度,相对于单独支持向量机(SVM)预测方法有较大提高,预测效果令人满意。

[参考文献]

[1] 刘 军,林叔忠,胡和平. 水源水库富营养化问题及生态治理途径 [J]. 水利渔业,2007,27(5):75.
Liu J,Lin S Z,Hu H P. Eutrophication problem and ecological controlling method of headwaters reservoirs [J]. Water Conservancy & Fishery,2007,27(5):75. (in Chinese)

[2] 金丹越,黄艳菊. 天津于桥水库主要环境问题及其防治对策

[J]. 环境科学研究,2004,17(增刊):77-85.
Jin D Y,Huang Y J. The environmental problems and controlling countermeasures of Tianjin Yuqiao reservoir [J]. Research of Environmental Sciences,2004,17(Suppl):77-85. (in Chinese)

[3] 杨 斌,翟雪梅,王 强. 长江河口南支水域溶解氧动力学模型的应用 [J]. 中国海洋大学学报,2004,34(2):253-259.
Yang B,Zhai X M,Wang Q. A 1-D oxygen hydrodynamic model for the southern part of the Yangtze estuary [J]. Journal of Ocean University of China,2004,34(2):253-259. (in Chinese)

[4] 罗 琳,李适宇,厉红梅. 夏季珠江口水域溶解氧的特征及影响因素 [J]. 中山大学学报:自然科学版,2005,44(6):118-121.
Luo L,Li S Y,Li H M. Characteristics of dissolved oxygen and its affecting factors in the Pearl River estuary in summer [J]. Journal of Zhongshan University: Natural Sciences Edition,2005,44(6):118-121. (in Chinese)

[5] 欧红香,郑 铭,田中雅史. 长良川水质中溶解氧和化学耗氧量预测模型的建立 [J]. 四川环境,2003,22(3):91-94.
Ou H X,Zheng M,Tanaka M. COD and DO predicting models of Nagara River [J]. Sichuan Environment,2003,22(3):91-94. (in Chinese)

[6] 陈丽华,马生全. 运用 BP 神经网络研究河水中溶解氧的影响因素 [J]. 兰州交通大学学报:自然科学版,2005,24(1):55-58.
Chen L H,Ma S Q. Study of influential factor of DO of river by BP neural network [J]. Journal of Lanzhou Jiaotong University: Natural Sciences Edition,2005,24(1):55-58. (in Chinese)

[7] 陈丽华. BP 网络应用于黄河水质浓度的预测研究 [J]. 西北民族学院学报:自然科学版,2003,24(1):39-43.
Chen L H. Based of three algorithms of BP neural networks to predict groundwater' DO concentration[J]. Journal of Northwest Minorities University: Natural Sciences Edition,2003,24(1):39-43. (in Chinese)

[8] 杨家宽,肖 波,刘年丰,等. WASP6 水质模型应用于汉江襄樊段水质模拟研究 [J]. 水资源保护,2005,21(4):8-10.
Yang J K,Xiao B,Liu N F,et al. Simulation on the water quali-

- ty for Xiangfan Reach of Hanjiang River using WASP6 [J]. Water Resources Protection, 2005, 21(4): 8-10. (in Chinese)
- [9] 王文圣, 丁 晶, 向红莲. 小波分析在水文学中的应用研究及展望 [J]. 水科学进展, 2002, 13(7): 515-520.
- Wang W S, Ding J, Xiang H L. Application and prospect of wavelet analysis in hydrology [J]. Advances in Water Science, 2002, 13(7): 515-520. (in Chinese)
- [10] 郭其一, 路向阳, 李维刚, 等. 基于小波分析和模糊神经网络的水文预测 [J]. 同济大学学报: 自然科学版, 2005, 33(1): 130-133.
- Guo Q Y, Lu X Y, Li W G, et al. Study on hydrological forecasting based on wavelet-fuzzy neural networks [J]. Journal of Tongji University: Natural Science Edition, 2005, 33(1): 130-133. (in Chinese)
- [11] 廖 杰, 王文圣, 李跃清, 等. 支持向量机及其在径流预测中的应用 [J]. 四川大学学报: 工程科学版, 2006, 38(6): 24-28.
- Liao J, Wang W S, Li Y Q, et al. Support vector machine method and its application to prediction of runoff [J]. Journal of Sichuan University: Engineering Science Edition, 2006, 38(6): 24-28. (in Chinese)
- [12] 王 炜, 赵利飞, 吴耿锋, 等. 地下水观测数据拟合与预测的支持向量机方法 [J]. 地震, 2007, 27(1): 9-15.
- Wang W, Zhao L F, Wu G F, et al. The support vector machine method for analogizing and forecasting groundwater data [J]. Earthquake, 2007, 27(1): 9-15. (in Chinese)
- [13] 王文圣, 丁 晶, 李跃清. 水文小波分析 [M]. 北京: 化学工业出版社, 2005.
- Wang W S, Ding J, Li Y Q. Hydrology wavelet analysis [M]. Beijing: Chemical Industry Press, 2005. (in Chinese)
- [14] Vapnik V N. An overview of statistical learning theory [J]. IEEE Trans Neural Network, 1999, 10(5): 988-999.
- [15] Suykens J A K, Vandewalle J. Recurrent least squares support vector machines [J]. IEEE Transactions on Circuits and Systems I, 2000, 47(7): 1109-1114.
- [16] Mallat S. A wavelet tour of signal processing [M]. 北京: 机械工业出版社, 2003.
- Mallat S. A wavelet tour of signal processing [M]. Beijing: Machinery Industry Press, 2003.
- [17] Suykens J A K, Vandewalle J. Least square support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [18] 姜谕男, 梁 冰. 基于 PSO-SVM 的大坝渗流监测时间序列非线性预报模型 [J]. 水利学报, 2006, 37(3): 331-335.
- Jiang A N, Liang B. Nonlinear time series prediction model for dam seepage flow based on PSO-SVM [J]. Journal of Water Conservancy, 2006, 37(3): 331-335. (in Chinese)
- [19] 赵洪波, 冯夏庭, 尹顺德. 基于支持向量机的岩体工程分级 [J]. 岩土力学, 2002, 23(6): 698-701.
- Zhao H B, Feng X T, Yin S D. Classification of engineering rock based on support vector machine [J]. Rock and Soil Mechanics, 2002, 23(6): 698-701. (in Chinese)
- [20] 张宏伟, 岳 琳, 王 亮. 基于径向基函数的城市日用水量预测方法 [J]. 天津大学学报, 2006, 39(4): 486-489.
- Zhang H W, Yue L, Wang L. Prediction method for based on radial basis function for daily water consumption [J]. Journal of Tianjin University, 2006, 39(4): 486-489. (in Chinese)
- [21] 刘 涛. 水库水质参数预测和富营养化预警 [D]. 天津: 天津大学, 2005.
- Liu T. Predictions on water quality parameters and advance warning on eutrophication in reservoir [D]. Tianjin: Tianjin University, 2005. (in Chinese)
- [22] 曲文龙, 樊广佳, 杨炳儒. 基于支持向量机的复杂时间序列预测研究 [J]. 计算机工程, 2005, 31(23): 1-3.
- Qu W L, Fan G Q, Yang B R. Research on complicated time series prediction based on support vector machines [J]. Computer Engineering, 2005, 31(23): 1-3. (in Chinese)