Journal of Water Resources Research 水资源研究, 2012, 1, 347-352
http://dx.doi.org/10.12677/jwrr.2012.15053 Published Online October 2012 (http://www.hanspub.org/journal/jwrr.html)

Study on Runoff and Sediment Variation of the Dongting Lake Based on PPR and SVM Model*

Shan Zhao[1,2], Nianqing Zhou[1,2], Zhengzui Li[3]

[1]Department of Hydraulic Engineering, Tongji University, Shanghai
[2]Key Laboratory of Yangtze River Water Environment, Shanghai
[3]Hydrology and Water Resources Bureau of Hunan Province, Changsha
Email: zhaoshan6310@126.com

Received: May 4th, 2012; revised: May 18th, 2012; accepted: May 28th, 2012

Abstract: The Dongting Lake is very important for flood storage and water sources in the midstream of the Yangtze River. Due to the dual effects of nature and human activities, several major changes have been taken place on the runoff and sediment conditions and the relationships between rivers and lakes successively. In order to explore the relationships of runoff and sediment in and out the lake, the existing hydrologic and sediment and other observation data of the Dongting Lake area are fully used, and two non-linear simulation models, Projection Pursuit Regression (PPR) and Support Vector Machine (SVM) are established. The simulation errors are also compared. The results show that the Support Vector Machine (SVM) has better validity and credibility, which could be used as an effective tool to simulate the complicated river system. This finding provides theoretical support and scientific basis for comprehensive improvement and ecological restoration of the Dongting Lake.

Keywords: Dongting Lake; Runoff; Sediment Variations; Projection Pursuit Regression; Support Vector Machine

基于 PPR 和 SVM 模型研究洞庭湖径流与输沙量变化*

赵　姗[1,2]，周念清[1,2]，李正最[3]

[1]同济大学水利工程系，上海
[2]长江水环境重点实验室，上海
[3]湖南省水文水资源勘测局，长沙
Email: zhaoshan6310@126.com

收稿日期：2012 年 5 月 4 日；修回日期：2012 年 5 月 18 日；录用日期：2012 年 5 月 28 日

摘　要：洞庭湖是长江中游典型的吞吐型调蓄湖泊。受自然和人为因素影响，其输入和输出的径流与输沙量关系先后发生了多次大的调整过程。为了探明洞庭湖水沙出入湖量变化和相互关系，本文利用现有水文泥沙等观测试验资料，基于投影寻踪回归(PPR)和支持向量机模型(SVM)对洞庭湖径流与输沙量变化特征进行了模拟和验证，并对模拟误差进行了对比。结果表明，两种模型均可以用于模拟洞庭湖水网径流与输沙量关系，但支持向量机模型(SVM)精确度较高，适用性较好，可为洞庭湖的综合整治提供理论支撑和科学依据。

关键词：洞庭湖；径流量；输沙量；投影寻踪回归(PPR)；支持向量机(SVM)

*基金项目：湿地演替带氧化还原电位及其对氮素迁移转化的影响(项目编号：20110072110020)。
作者简介：赵姗(1987-)，女(汉)，河北省保定人，博士研究生，主要从事水文水资源工作。

1. 研究背景

洞庭湖一直以来是众多水文学者研究的热点问题[1-3]。随着三峡工程的建设与运行，洞庭湖防洪功能已显著提高[4,5]，但与此同时其径流与输沙量关系也发生了很大变化[6,7]。洞庭湖水系与珠江三角洲河网等相比，来水来沙情况更为复杂多变，为了探明洞庭湖水沙出入湖量变化和相互关系，选择合适的模拟方法非常重要。

针对复杂水系径流与输沙量关系的研究很多，也提出了很多模型[8]。投影寻踪(Projection Pursuit)是一种用来分析和处理高维观测数据，尤其是非线性、非正态高维数据的新兴方法[9,10]。1981 年，Freidman 和 Stuetzle[11]基于投影寻踪的思想最先给出了投影寻踪回归方法，其主要目的在于解决高维空间中的回归问题。王顺久等[12]运用投影寻踪评价模型对全国区域水资源承载能力和淮河流域水资源承载能力 2 个实例进行了分析；刘卫林[13]运用投影寻踪回归模型评价了地下水环境的脆弱性。支持向量机(Support Vector Machines)是在统计学理论基础上发展起来的一种新的机器学习方法[14,15]。支持向量机在最小化样本点误差的同时，缩小模型的复杂度，提高了模型的泛化能力。李正最[16,17]等首次将支持向量机理论引入洞庭湖水量交换和水沙模拟研究中，初步建立了基于支持向量机的洞庭湖水沙交互模型，并得到了较好的模拟结果。

目前投影寻踪回归(PPR)和支持向量机(SVM)模型均已成功运用到多个专业领域，但很少有人将这两种模型进行比较。本文根据洞庭湖 1956~2008 年实测水文资料，将洞庭湖径流与输沙量关系看作是一种多路水沙交互作用的复杂的小样本和非线性问题，基于 PPR 和 SVM 模型对洞庭湖径流与输沙量关系进行了模拟与验证，并对模拟误差进行了对比。

2. 投影寻踪回归(PPR)和支持向量机(SVM)原理

2.1. 投影寻踪回归(PPR)

投影寻踪的基本思想是：利用计算机技术，把高维数据通过某种组合投影到低维子空间上，并通过极小化某个投影指标，寻找出能反映原数据结构或特征的投影，以达到研究和分析高维数据的目的。投影寻踪模型如下：设 $y = f(X)$ 和 $X = (x_1, x_2, \cdots x_p)$ 分别为一维和 p 维随机变量，为了客观反映高维非线性结构特征，投影寻踪回归采用一系列岭函数的“和”去逼近回归函数，即：

$$f(X) \approx \sum_{m=1}^{M} G_m(Z_m) = \sum_{m=1}^{M} G_m\left(a_m^T X\right) = \sum_{m=1}^{M} G_m\left(\sum_{j=1}^{p} a_{mj} x_j\right) \tag{1}$$

式中：$G_m(Z_m)$ 为第 m 个岭函数；M 为岭函数的个数；$Z_m = a_m^T X$ 为岭函数的自变量，它是 p 维随机变量 X 在 a_m 方向上的投影；a_m 为投影方向。

投影寻踪回归模型仍采用最小二乘法作为极小化判别准则，即以式(1)中的参数 a_{mj}、G_m 和岭函数个数 M 的适当组合，使下式

$$L = E\left[y - \sum_{m=1}^{M} G_m\left(\sum_{j=1}^{p} a_{mj} x_j\right)^2\right] \tag{2}$$

达到极小。对于式(2)中的非线性系统模型，实现投影寻踪回归的步骤如下：

1) 确定岭函数的个数 M。

2) 选择 M 个彼此正交的投影方向 $a_1, a_2, \cdots, a_m$，建立初步回归模型：

$$f(X) = \sum_{m=1}^{M} \sum_{i-0}^{r} C_{mi} h_{mi}\left(a_m^T X\right) \tag{3}$$

3) 分组优化。即将 $a_{mj}(j = 1,2,\cdots,p)$和 G_m(即 $h_{mi}(i = 0,1,\cdots,r)$)划为一组，$m = 1,2,\cdots,M$，共有 M 组。除去其中一组外，对另外的 $M-1$ 组用 2)中得到的值作为初值，对留下的一组参数寻优。求得结果后把这一组参数的极值点作为初值，另选一组参数寻优，反复多次直到最后选取的一组参数值，使式(2)不再减小为止。

4) 参数处理并输出回归模型：

$$f(X) = \sum_{m=1}^{M} \sum_{i-0}^{r} C_{mi} H_{mi}\left(\hat{a}_m^T X\right) \tag{4}$$

2.2. 支持向量机(SVM)回归

支持向量机的基本思想是用少数支持向量代表整个样本集，本质上是通过某一事先选择好的非线性函数 $\varphi(\bullet)$ 将训练集数据 X 映射到一个高维线性特征空间 H，在这个维数可能为无穷大的线性空间中按结

构风险最小化原理构造最优分类面。并利用原空间的核函数取代高维特征空间 ω 和 $\Phi(x)$ 的点积运算，从而避免了复杂的点积计算。对于给定的样本数据集 $\{(x_i, y_i)|i=1,2,\cdots,l\}$，其中 x_i 为输入值，y_i 为预测值。要求拟合的函数形式为：

$$f(x)=w\varphi(x)+b \tag{5}$$

SVM 用来估计回归函数时，常分为线性和非线性拟合回归两类。由上式可求得线性回归函数为

$$f(x)=wx+b=\sum_{SV}(\alpha_i-\alpha_i^*)(x_i-x)+b \tag{6}$$

对于非线性的情况，引入核函数即可。此时求得的是非线性回归函数为：

$$f(x)=w\varphi(x)+b=\sum_{SV}(\alpha_i-\alpha_i^*)K(x_i-x)+b \tag{7}$$

其中 $K(x,x_i)=\varphi(x)\varphi(x_i)$ 称为核函数，其必须满足 Mercer 条件。常见的核函数有多项式核函数、径向基核函数和 Sigmoid 核函数，三种核函数中都有参数σ。

目前最常用的支持向量机为Suykens于1999年提出的改进的最小二乘支持向量机，采用二次规划方法代替传统的支持向量机来解决函数估计问题，其利用结构风险原则时，在优化目标中选取了不同的损失函数。核函数的参数σ和最小二乘支持向量机的参数取值对模型的推广预测能力有很大的影响，若取值不当，均会增大模型误差，其取值通常采用试算法或经验法[18,19]，本文采用混沌优化算法进行参数寻优。

3. 洞庭湖径流与输沙量关系模拟

3.1. 数据来源和研究区概况

水文数据来源于 1956~2008 年洞庭湖流域水文年鉴和主要水文站监测资料。洞庭湖位于湖南北部、长江荆江南岸，跨越湘鄂两省。北面有松滋、太平、藕池和调弦口(于 1958 年封堵)，分泻长江水沙，南有湘、资、沅、澧四水汇入，周边汨罗江、新墙河等中小河流直接入湖，经洞庭湖调蓄，于城陵矶汇入长江，是长江中下游重要的调蓄型湖泊，对分泻荆江洪水和保障下游径流供给起着十分显著的作用(图 1)。新中国成立以来，长江中游河段经历了调弦口封堵、下荆江系统裁弯取直、葛洲坝和三峡水库建成发电等；湖南省湘、资、沅、澧四水流域包括柘溪、五强溪等干流骨干性工程在内的 13,000 多座各种水利工程和水土保持工程，但是很多工程并没有取得预期效果，如下荆江裁弯工程[20]。

3.2. 模型整体结构

洞庭湖水沙系统具有十分明显的非线性特征，因此在建模的具体手段上分别选用投影寻踪回归和支持向量机两种方法。用 1956~2004 年洞庭湖区水沙序列进行模型拟合，以 2005~2008 年洞庭湖区水沙序列进行模型检验。洞庭湖出口城陵矶站的径流量和输沙量可简单地表述为以下非线性结构，即：

$$Q_d=\varphi(Q_u,Q_\lambda,q,V) \tag{8}$$

$$S_d=\psi(Q_u,S_u,Q_\lambda,S_\lambda,q,Q_d,\chi_{地形},\cdots) \tag{9}$$

式中：Q_d 为城陵矶出口断面的径流量；S_d 为城陵矶出口断面的泥沙；Q_u 为四水入流量；S_u 为四水来沙量；Q_λ 为三口分流量；S_λ 为三口分沙量；q 为区间产水量；V 为洞庭湖调蓄量；$\chi_{地形}$ 为洞庭湖区地形特征；$\varphi(\bullet)$ 为水量交换作用函数；$\psi(\bullet)$ 为水沙交互作用函数。

3.3. 模拟与预测误差

按照建模序列和检验序列，分别统计两种模型的最大误差和绝对平均误差。因检验序列过短不宜独立计算误差标准差，故按建模序列和检验序列合并计算。主要误差指标计算公式如下：

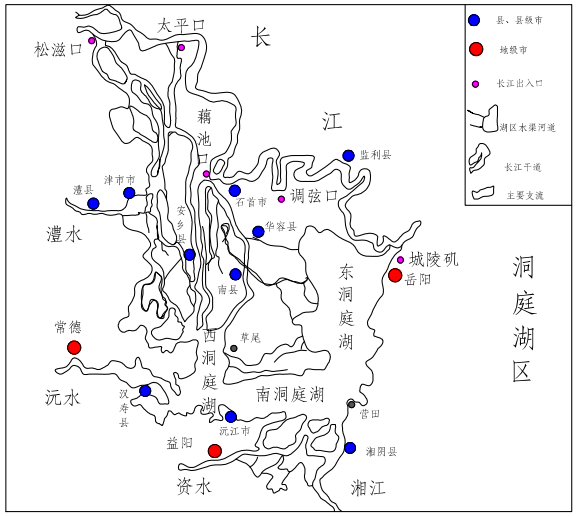


Figure1. Stream structure of Dongting Lake
图 1. 洞庭湖区河网网水系结构图

$$e_i = \left(y_i - \hat{y}_i\right)/\hat{y}_i \times 100\% \quad (10)$$

$$e_{\max} = \max_{i=1}^{n}\left[ABS\left(e_i\right)\right] \quad (11)$$

$$e_{\text{mean}} = \sum_{i=1}^{n}\left[ABS\left(e_i\right)\right]/n \quad (12)$$

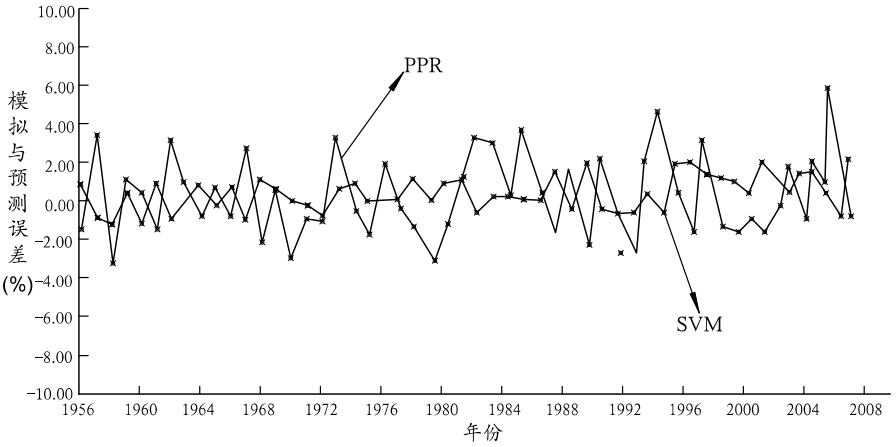
$$S_e = \sqrt{\sum_{i=1}^{n}\left(e_i\right)^2/\left(n-1\right)} \quad (13)$$

式中：e_i 为第 i 个样本的拟合(预报)误差；y_i 为第 i 个样本实测值；$\hat{y}_i$ 为第 i 个样本拟合或预测值；$e_{\max}$ 为最大拟合或预报误差；$\max(\bullet)$ 为取大运算符；$ABS(\bullet)$ 为绝对值运算符；e_{mean} 为平均绝对误差；S_e 为误差标准差；n 为样本总数。对上述建立的两种模型分别进行回顾检验和外推预报，以式(10)计算相对误差，误差分布情况见图 2。

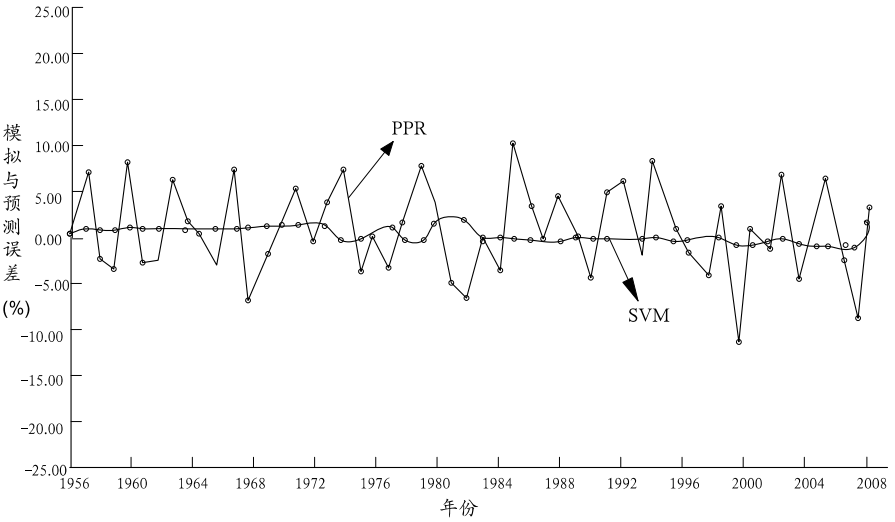
对于所建立的 2 种模型，用城陵矶站年径流和年输沙量按式(10)~(13)统计误差，计算结果见图 3。

3.4. 结果分析

从图 2 和图 3 可以看出，所建立的两种模型均具备一定的复杂系统仿真能力。而就模型的类别而言，以 SVM 模型的精度较高，PPR 略低；就模型的输出物理量而言，两种模型的径流量模拟输出精度均高于



(a)



(b)

Figure 2. Error analysis of simulation and prediction of runoff and sediment relationship. (a) Run-off; (b) Sediment
图 2. 洞庭湖径流与输沙量关系模型拟合与预报误差分析。(a) 径流量；(b) 输沙量

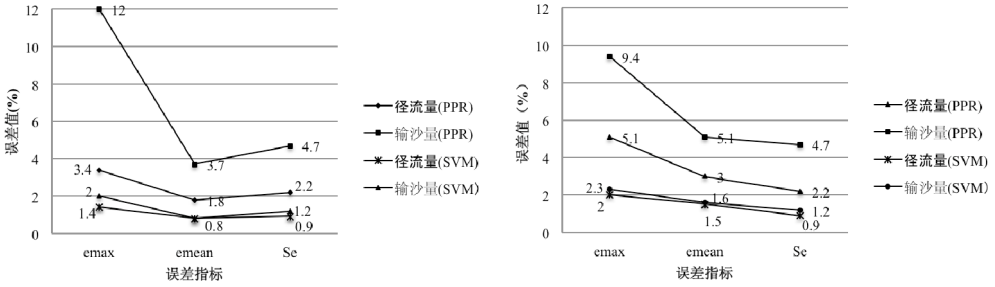


Figure 3. Error statics of runoff and sediment relationship of Dongting Lake. (a) Modeling series; (b) Test series
图 3. 洞庭湖径流与输沙量关系模拟与检验误差统计。(a) 建模序列；(b) 检验序列

输沙量，说明江湖水沙演化中输沙量的影响因素更为广泛，演化机制更为复杂，非线性特征更为显著；就模型的推广和泛化能力而言，PPR 检验序列精度对建模序列精度有所下降，SVM 检验序列精度基本与建模序列相匹配，没有表现出明显的下降趋势。可见 SVM 的有效性和可信性较好，其系统数据与模型数据之间具有较好的一致性，因而其对复杂水网水沙交互作用的拟合和推广能力较强。因此，运用 SVM 模型模拟计算的洞庭湖水沙出入湖量和区域泥沙淤积情况，可供江湖治理决策参考。

4. 结语

本文利用洞庭湖近 50 年来的水文观测资料，基于投影寻踪回归和支持向量机分别建立了洞庭湖径流与输沙量两种非线性仿真模型，得到以下结论：

1) 通过两种模型的误差比较，SVM 模型的精度较高，说明 SVM 模拟和预测的结果与实测值吻合度高，试用、可操作性强，为复杂水网区的水沙分析提供了一种新方法。

2) 支持向量机的推广性能与模型的参数选择有很大关系。因此，如何根据训练样本选择合适的模型参数，以保证建立好的模型有很好的推广性能，成为设计支持向量机关键一步。

3) 通过模拟可以看出，两种模型中洞庭湖的径流量输出精度均高于输沙量，说明洞庭湖输沙量变化涉及因素更多，而不仅仅与径流量有关。影响输沙量因素，有待下一步研究。

参考文献 (References)

[1] 李景保. 洞庭湖区 1996 年特大洪涝灾害的特点与成因分析[J]. 地理学报, 1998, 53(2): 166-173.
LI Jingbao. A study on the features and causes of the flood disasters in Dongting Lake plain in 1996. Acta Geographica Sinica, 1998, 53(2): 166-173. (in Chinese)

[2] 童潜明. 三峡水库运行后对洞庭湖防洪和生态的思考[J]. 国土资源科技管理, 2001, 3: 1-6.
TONG Qianming. Management of land reflections on flood control and ecology of Dongting Lake after operation of the three gorges reservoir. Scientific and Technological Management of Land and Resources, 2001, 3: 1-6. (in Chinese)

[3] 甘明辉, 刘卡波, 杨大文, 等. 洞庭湖四口河系防洪、水资源与水环境研究[J]. 2011, 30(5): 5-9.
GAN Minghui, LIU Kabo, YANG Dawen, et al. Analysis on the flood control, water supply and water environment projection in the region of Dongting Lake along Yangtze River. Journal of Hydroelectric Engineering, 2011, 30(5): 5-9. (in Chinese)

[4] 穆锦斌, 张小峰. 荆江－洞庭湖水沙变化影响分析[J]. 水利水运工程学报, 2011, 1: 84-91.
MU Jinbin, ZHANG Xiaofeng. Analysis on water-sediment transportation variation of Jingjiang River-Dongting Lake. Hydro-Science and Engineering, 2011, 1: 84-91. (in Chinese)

[5] 李景保, 代勇, 欧朝敏, 等. 长江三峡水库蓄水运用对洞庭湖水沙特性的影响[J]. 水土保持学报, 2011, 25(3): 215-219.
LI Jingbao, DAI Yong, OU Chaomin, et al. Effects of store water application of the Three-Gorges Reservoir on Yangtze River on water and sediment characteristics in the Dongting Lake. Journal of Soil and Water Conservation, 2011, 25(3): 215-219. (in Chinese)

[6] 马元旭, 来红州. 荆江与洞庭湖区近 50 年水沙变化的研究[J]. 水土保持研究, 2005, 12(4): 103-106.
MA Yuanxu, LAI Hongzhou. Research on the variations of the water and sediment for recent 50 years in the Jingjiang River and Dongting Lake Area. Research of Soil and Water Conservation, 2005, 12(4): 103-106. (in Chinese)

[7] 毛北平, 梅军亚, 张金辉, 等. 洞庭湖三口洪道水沙输移变化分析[J]. 人民长江, 2010, 2: 38-42.
MAO Beiping, MEI Junya, ZHANG Jinhui, et al. Analysis of water and sediments transportation of three river-outlets flood chnnels from Yangtze River to Dongting Lake. Yangtze River, 2010, 2: 38-42. (in Chinese)

[8] 吴作平, 杨国录, 甘明辉. 荆江－洞庭湖水沙模型研究[J]. 水利学报, 2003, 7: 96-100.
WU Zuoping, YANG Guolu and GAN Minghui. Mathematical model for flow-sediment transportation of Jingjiang River-Dongting Lake network. Journal of Hydraulic Engineering, 2003, 7: 96-100. (in Chinese)

[9] FRIEDMAN, J. H., TUKEY, J. W. A projection pursuit algorithm or exploratory data analysis. IEEE Transactions on Computer, 1974, 23(9): 881-890.

[10] KRUSCAL, J. B. Toward a practical method which helps uncover the structure of a set of multivariate observations by find-

ing the linear transformation which optimizes a new index of condensations. Statistical Computer, New York: Academic, 1969.

[11] FREIMAN, J. H., STUEZLE, W. Projection pursuit regression. Journal of American Statistic Association, 1989, 76: 817-823.

[12] 王顺久, 侯玉, 张欣莉, 等. 流域水资源承载能力的综合评价方法[J]. 水利学报, 2003, 1: 88-92.
WANG Shunjiu, HOU Yu, ZHANG Xinli, et al. Comprehensive evaluation method for water resources carrying capacity in river basins. Journal of Hydraulic engineering, 2003, 1: 88-92. (in Chinese)

[13] 刘卫林. 遗传投影寻踪回归模型在地下水环境脆弱性评价中的应用[J]. 南昌工程学院, 2011, 30(3): 1-5.
LIU Weilin. Application of projection pursuit model based on genetic algorithm to evaluation of the groundwater environmental vulnerability. Journal of Nangchang Institute of Technology, 2011, 30(3): 1-5. (in Chinese)

[14] VAPNIK, V. N. The nature of statistic learning theory. New York: Spring Verlag, 1995.

[15] ZHANG, X. G. On statistical learning theory and support vector machine. Acta Automatica Sinica, 2000, 26(1): 32-42.

[16] 李正最, 谢悦波, 徐冬梅. 基于支持向量机的洞庭湖水量交换模型[J]. 水电能源科学, 2009, 27(5): 18-20.
LI Zhegnzui, XIE Yuebo and XU Dongmei. Water exchange model in Dongting Lake based on support vector machine. Water Resources and Power, 2009, 27(5): 18-20. (in Chinese)

[17] 李正最, 谢悦波. 基于支持向量机的洞庭湖区域水沙模拟[J]. 水文, 2010, 30(2): 44-49.
LI Zhegnzui, XIE Yuebo. Simulation of water and sediment in Dongting Lake based on support vector machine. Journal of China Hydrology, 2010, 30(2): 44-49. (in Chinese)

[18] 田盛丰, 黄厚宽. 基于支持向量机的数据库学习算法[J]. 计算机研究与发展, 2000, 37(1): 17-22.
TIAN Shengfeng, HUANG Houkuan. Database learning algorithms based on supportive machine. Journal of Computer Research & Development, 2000, 37(1): 17-22. (in Chinese)

[19] 苏高利, 邓芳萍. 关于支持向量回归机的模型选择[J]. 科技通报, 2006, 22(2): 154-158.
SU Gaoli, DENG Fangping. Introduction to model selection of SVM regression. Bulletin of Science and Technology, 2006, 22(2): 154-158. (in Chinese)

[20] 潘庆燊. 下荆江人工裁弯 30 年[J]. 人民长江, 2001, 32(5): 27-29.
PAN Qingshen. Over 30 years about the artificial cut-off project on the lower Jingjiang River. Yangtze River, 2001, 32(5): 27-29. (in Chinese)