

Research on Medium and Long Term Runoff Forecast in Yangtze River Basin

Junwei Jia^{1,2}, Liping Zhang^{1,2}, Lian Liu^{1,2}, Lijie Shan^{1,2}

¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China

²Hubei Collaborative Innovation Center for Water Resources Security, Wuhan, China
Email: 1240756256@qq.com

Received: Jul. 14th, 2014; revised: Jul. 21st, 2014; accepted: Jul. 29th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Yangtze River is one of China's most economically developed regions, and the quantity of water directly affects the economic and social development in the region, so an accurate medium and long-term forecast is significant for multi-reservoir scheduling, water resources allocation and rational utilization. In this paper, Pingshan, Yichang, Datong and Hankou four stations in the Yangtze River basin were selected as the research objects, based on 75 predictors, including 74 atmospheric circulation index and pre-runoff, using the correlation coefficient method to preliminary select the predictors and stepwise regression method to optimize the predictors, a runoff forecasting model of monthly scale and ten-day scale based on the Support vector machine (SVM) was established. And the applicability of the model in the Yangtze River basin was quantitatively analyzed. In addition, the prediction precision of the model was compared with that of artificial neural network prediction model. The results indicate that the model based on SVM can't meet the actual application requirements, because the average qualified rates during the test period were only respectively 49.29% and 54.49% in the medium and long-term runoff forecasting of monthly scale and ten-day scale. But the monthly runoff results calculated by ten-day scale are superior to those by monthly scale. The former can provide a reference for the work of the Yangtze River hydrological forecasting. Relatively speaking, the prediction precision of the model based on Support vector machine (SVM) is better than that of artificial neural network model.

Keywords

Stepwise Regression, Support Vector Machine, Yangtze River Basin, Medium and Long-Term Runoff Forecast

作者简介: 贾军伟(1991-), 男, 硕士研究生, 研究方向为水文预报研究。

长江干流站中长期径流预报方法研究

贾军伟^{1,2}, 张利平^{1,2}, 刘 恋^{1,2}, 闪丽洁^{1,2}

¹武汉大学水资源与水电工程国家重点实验室, 武汉

²水资源安全保障湖北省协同创新中心, 武汉

Email: 1240756256@qq.com

收稿日期: 2014年7月14日; 修回日期: 2014年7月21日; 录用日期: 2014年7月29日

摘 要

长江流域是我国经济最发达的地区之一, 水资源量的多少直接影响着该地区经济社会的发展, 因此准确的中长期预报对于水库群联合调度、水资源调配和合理利用具有重要的意义。本文以长江流域屏山、宜昌、大通和汉口4个站点为研究对象, 依据74项大气环流指数和前期径流共75项预报因子, 采用相关系数法初选及逐步回归法优选预报因子, 建立了基于支持向量机的月、旬径流预报模型, 定量分析了模型在长江流域的适用性, 并与人工神经网络预测模型进行比较, 结果表明该模型在长江流域中长期径流月、旬预报中, 检验期平均合格率分别为49.29%和54.49%, 达不到实际应用的需求, 而旬尺度计算月径流的结果优于月尺度的模拟结果, 有较好的预报精度, 可为长江流域水文预报工作提供参考。相对而言, 支持向量机模型的预测精度优于人工神经网络模型。

关键词

逐步回归, 支持向量机, 长江流域, 中长期径流预报

1. 引言

中长期径流预报在防汛、抗旱、供水、发电、养殖、旅游、航运及改善生态环境等方面发挥显著作用[1]。同时中长期径流预报系统的建立可为水利工程年和月水资源计划调度的制定提供年、月来水预报信息, 能使人们在协调防洪、兴利、生态之间及各部门用水之间矛盾时, 及早采取措施进行统筹安排, 以获得最佳综合效益。传统的预报方法主要有成因分析和水文统计方法, 成因分析法是基于分析大气环流与水文要素之间的联系, 水文统计法是通过水文资料的统计分析进行概率预测, 它包括两大类, 一类是分析水文要素自身随时间变化的统计规律, 并以此预测, 如历史演变法、时间序列分析法等; 另一类是用多元线性回归法或逐步线性回归法, 建立预报方案进行预报[2]。20世纪70年代后期Box[3]等人提出的自回归滑动平均(autoregressive moving-average, ARMA)模型在水文预报中得到广泛应用, 90年代以来, 人工神经网络(ANN)[4]在水文预报中的应用逐渐增多。近年来, 基于新的数学分支和计算机技术的发展, 中长期水文预报拓展了新的研究途径, 支持向量机由于其预测精度较高, 鲁棒性强, 泛化能力强等诸多优点, 在处理小样本数据时具有其他模型无法比拟的优势, 且具有良好的拟合和外推能力[5], 越来越多地被应用到水文预报中。中长期径流预报的一个重要环节是预报因子的合理识别, 为预报模型的构建奠定必要的基础, 不同的识别方法会有不同的预报因子, 进而导致不同的预报效果。本文以长江流域1980~2012共33年径流资料为依据, 采用基于逐步回归-支持向量机方法对长江流域屏山、宜昌、大通、汉口4个站点未来的逐月和逐旬径流进行预报, 并对各模型预报结果的优劣比较分析。

2. 支持向量机模型原理

支持向量机具有坚实的数学理论基础,是专门针对小样本学习问题提出的,通过统计学习中的VC维理论和结构风险最小化原则实现[6]。它将具体问题转化为一个二次型寻优问题,理论上可以得到全局最优解,并且对于小样本学习具有良好的泛化性能,能够进行高效、精确的预测,较好地解决了传统方法的经验依赖、局部极值等问题[7]。

支持向量机通过实现选择好的某一个非线性变换,将输入向量 x 映射到高维特征空间 Z ; 在这一特征空间中,构造一个最优分类超平面,使分类边界,即分类平面与最近点(支持向量)之间的距离最大,并且这种非线性变换是通过定义合适的核函数来实现,然后将 SVM 问题转化为一个二次规划问题,从而求解。

给定训练集

$$T = \{(x_i, y_i), \dots, (x_l, y_l)\} \in (R^n \times R) \quad (1)$$

式中: $x_i \in R^n, y_i \in R, i=1,2,\dots,l$, x_i 为输入变量, y_i 是 x_i 对应的输出值, l 为样本个数。

支持向量机回归的基本思想就是通过一个非线性映射 ϕ 将数据 x_i 映射到高维特征空间 F , 并在这个空间进行线性回归, 即

$$f(x) = w^T \phi(x) + b \quad (2)$$

式中: w 为超平面的权值向量, b 为偏置项。

支持向量机实际上就是在约束条件(4)下求解下面的优化问题:

$$\min_{w,b,\xi,\xi^+} J = \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi + \xi^+) \quad (3)$$

约束条件:

$$\text{S.t.} \begin{cases} y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* & i=1,2,\dots,l \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4)$$

其中 ξ_i, ξ_i^* 为松弛变量, 分别表示在误差 ε 约束下 $(y_i - [w^T \phi(x_i) + b]) < \varepsilon$ 的训练误差的上限和下限; ε 为 Vapnik- ε 不敏感代价函数所定义的误差, 当预测值在定义的误差 ε 内, 代价函数为 0; 当预测值在定义的误差 ε 外时, 代价函数为预测值与误差 ε 差的幅值。常数 $C > 0$, 为惩罚函数, 它控制对超出误差 ε 的样本的惩罚程度。

为求解这样一个优化问题, 根据 Kuhn-Tucker 条件, 引入拉格朗日函数:

$$L = \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi + \xi^*) - \sum_{i=1}^l a_i (\xi_i + \varepsilon - y_i + w \cdot x_i + b) - \sum_{i=1}^l a_i^* (\xi_i^* + \varepsilon + y_i - w \cdot x_i + b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (5)$$

其中 $a_i, a_i^*, \eta_i, \eta_i^* > 0$ 为拉格朗日乘子, 根据 KKT 条件, 如下等式和约束条件成立:

$$\text{S.t.} \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i^* \leq C & i=1,2,\dots,l \\ w = \sum_{i=1}^l (a_i - a_i^*) \phi(x_i) \end{cases} \quad (6)$$

因此，非线性回归问题可以通过解式(3)的对偶问题求解：

$$\max_{a, a^*} W(a_i, a_i^*) = -\frac{1}{2} \sum_{i,j=1}^l (a_i^* - a_i)(a_j^* - a_j) k(x_i, x_j) - \varepsilon \sum_{i=1}^l (a_i + a_i^*) + \sum_{i=1}^l y_i (a_i - a_i^*) \quad (7)$$

其中核函数 $K(x_i, x_j) = \varphi(x_i)^T \cdot \varphi(x_j)$ 描述了高维特征空间的内积，可以在满足 Mercer 条件的情况下选取。常见的核函数有线性核函数、多项式核函数、径向基核函数、sigmoid 核函数等，本文采用径向基核函数。

求解后得到 a_i 和 a_i^* 代入式(7)，并由式(2)得到回归函数：

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) K(x, x_i) + b \quad (8)$$

式(8)中 N 为支持向量个数。

3. 试验流域与耦合模型的建立

3.1. 试验流域概况

长江流域面积辽阔，约占全国陆地面积的 1/5，人口约占全国的 1/3，经济总量占全国的 41.1%。长江流域是我国经济最发达的地区之一，同时也是我国洪涝灾害最为严重的地区。长江流域相对丰沛的水资源条件为流域内农业、工业和社会发展提供了有力保障，在我国的社会发展过程中扮演者重要的作用。然而，受地理位置、气候及下垫面等因素的影响，长江流域降水、径流的地区分布不均，年内变化显著，是洪涝发生的重灾区。

3.2. 耦合模型的建立

通过分析长江流域屏山、宜昌、汉口、大通站断面的水文情势变化规律，识别影响该区域的大气环流因素，对该 4 个站点进行月、旬径流预报。影响月、旬径流的因素是在 74 项大气环流指数和前期径流共 75 项预报因子中进行选取，而现行的统计预报模式中，对预报因子的优化选择的方法有很多，本文通过相关系数法[8]进行初选，然后利用逐步回归法[9]，优选出与月、旬径流相关程度较高的 6~8 个预报因子。预报因子需要进行线性归一化处理，使得支持向量机的样本数据在(0,1)范围内。通过样本训练模型，然后对于训练好的模型预测，给定输入数据，就可以得到相应的预测输出。支持向量机进行模型率定时，采用交叉验证的方法优化函数参数。

根据资料情况，将研究过程分为两个阶段，即率定期为 1981~2000 年，检验期为 2001~2012 年。

4. 结果分析

选取平均相对误差(ART)、平均合格率(AP)作为中长期径流预报结果的评价指标，评价模型的模拟能力。其计算公式如下：

$$ART = \frac{|Q_{模} - Q_{实}|}{Q_{实}} \quad (9)$$

$$AP = \frac{\sum_{i=1}^N P_i}{12} \times 100\% \quad (10)$$

其中 $Q_{实}$ 为实测值， $Q_{模}$ 为模拟值， p_i 指第 i 个月的合格率， $p_i = \frac{P_i}{P}$ ， P_i 是合格的结果数， P 是实测序

列的长度。而合格检验标准为： $C = (Y_{\max} - Y_{\min}) \times 20\%$ ， C 为误差限，当 $C > \text{abs}(E_i)$ ，即误差绝对值小于误差限时，认为该结果合格。反之，认为不合格。

由表 1 可以看出，四个站点率定期的月径流模拟平均相对误差为 3.62%，检验期为 23.77%，率定期的月径流模拟平均合格率达 96.77%，检验期则只有 49.29%，效果不理想。

由表 2 可以看出，四个站点旬径流模拟的相对误差在率定期和检验期与月径流模拟结果相差不大，但平均合格率在率定期和检验期都要优于月径流的模拟结果，特别是在检验期旬径流模拟的平均合格率相对于月径流模拟有了一定的提高，但预报效果仍然达不到实际应用的需求。

基于此，本文尝试先进行旬尺度的径流预报，然后将每个月 3 个旬的径流预报结果累加成月径流，结果如表 3 和图 1~图 4 所示。

表 3 及表 4 分别是由支持向量机与人工神经网络旬尺度计算所得的月径流结果，即根据旬径流模拟计算月径流的过程。对比表 1 和表 3 可以看出，由旬尺度计算出所得的月径流结果均优于月尺度模拟结果，且有了较大的提高。其中平均相对误差降低了 30% 左右，平均合格率同样提高了 35% 左右，大通站更是提高了 85.51%。比较表 3 和表 4 可以看出，支持向量机预测方法所得结果优于人工神经网络预测方法，其中检验期平均相对误差降低了约 10%，检验期平均合格率提高了 7%，尤其是汉口站和大通站均提高了 20% 左右，这是因为支持向量机模型以结构风险最小化原理为基础，对未来的径流样本有较好的泛化能力，其在解决小样本、非线性问题中表现出许多特有的优势。由图 1~图 4 可以看出，在检验期的 4

Table 1. The monthly runoff simulation results in each station
表 1. 各站点月径流模拟结果

站点	评价指标	屏山	宜昌	汉口	大通	平均
率定期	ART (%)	3.28	3.86	4.07	3.27	3.62
	AP (%)	97.08	97.92	95.42	96.67	96.77
检验期	ART (%)	21.15	23.82	22.39	27.70	23.77
	AP (%)	49.24	47.92	52.08	47.92	49.29

Table 2. The ten-day runoff simulation results in each station
表 2. 各站点旬径流模拟结果

站点	评价指标	屏山	宜昌	汉口	大通	平均
率定期	ART (%)	2.98	2.71	4.21	5.05	3.74
	AP (%)	98.89	98.61	96.11	95.14	97.19
检验期	ART (%)	21.00	25.64	23.54	26.05	24.06
	AP (%)	57.07	50.93	54.17	55.79	54.49

Table 3. The monthly runoff calculation result by ten-day scale
表 3. 由旬尺度计算月径流结果

站点	评价指标	屏山	宜昌	汉口	大通	平均
率定期	ART (%)	2.46	3.65	3.76	5.33	3.80
	AP (%)	99.58	97.92	98.75	95.42	97.92
检验期	ART (%)	17.53	20.66	18.64	9.80	16.66
	AP (%)	62.12	54.17	61.11	88.89	66.57

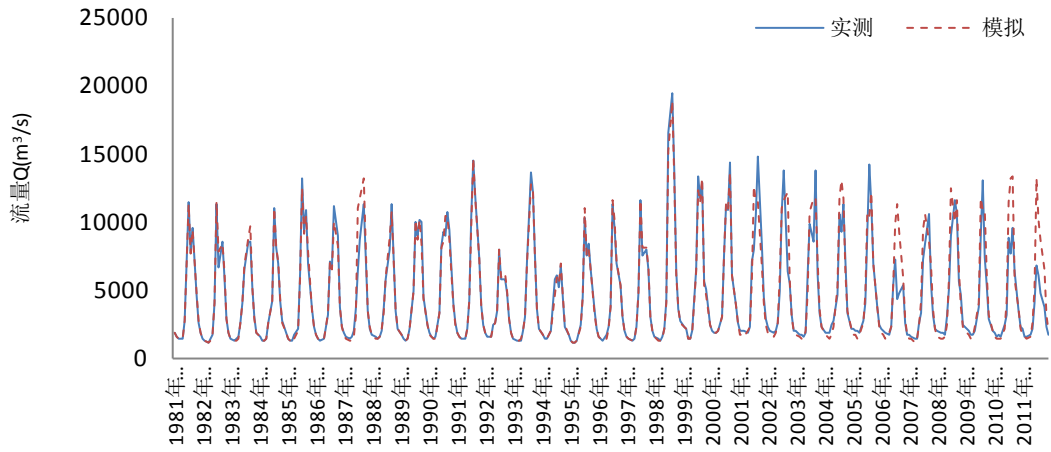


Figure 1. The monthly runoff simulation results calculated by ten-day scale in the Pingshan station
图 1. 屏山站旬尺度计算月径流模拟结果

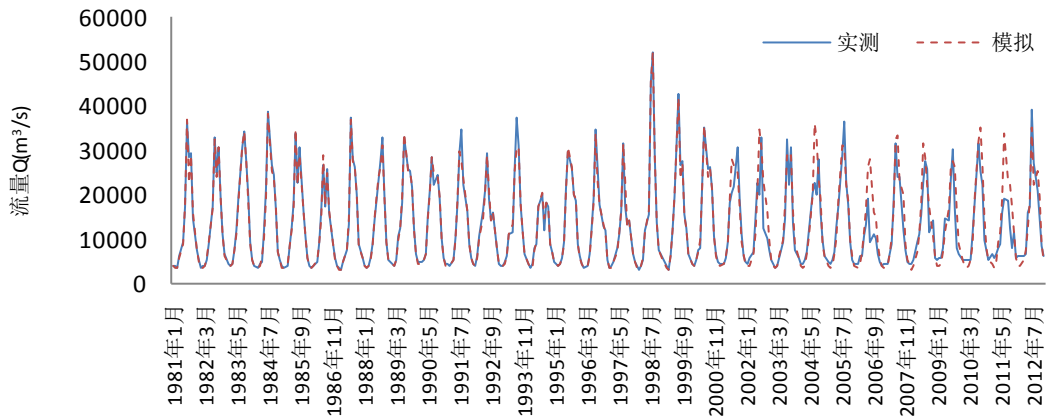


Figure 2. The monthly runoff simulation results calculated by ten-day scale in the Yichang station
图 2. 宜昌站旬尺度计算月径流模拟结果

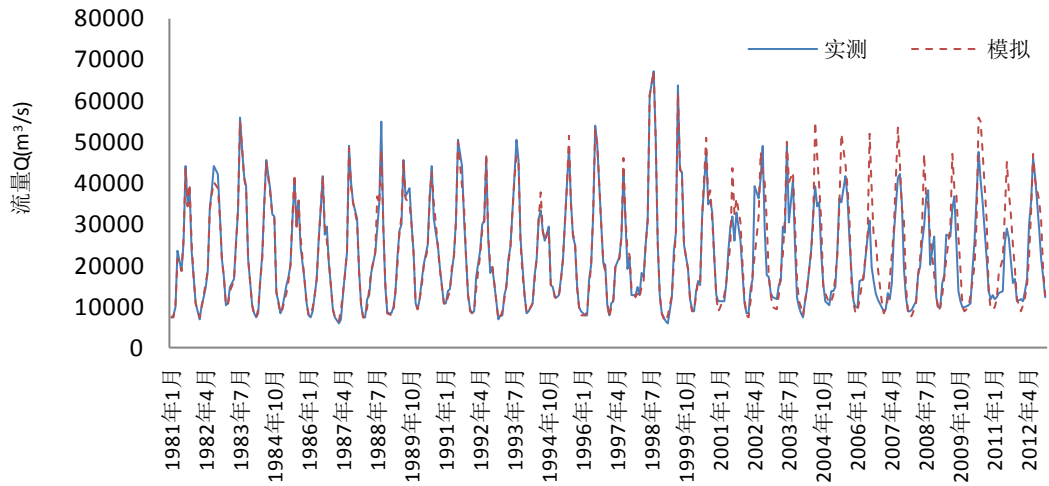


Figure 3. The monthly runoff simulation results calculated by ten-day scale in the Hankou station
图 3. 汉口站旬尺度计算月径流模拟结果

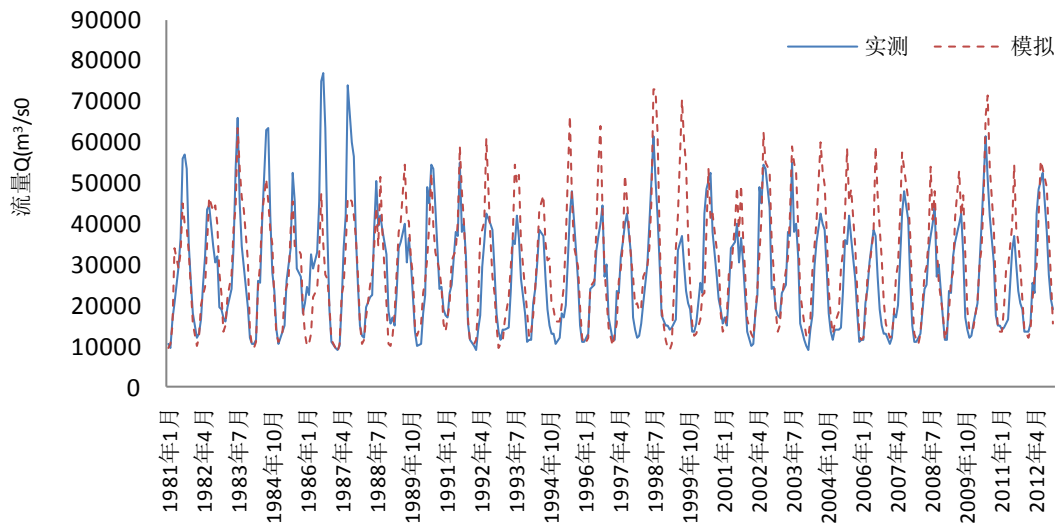


Figure 4. The monthly runoff simulation results calculated by ten-day scale in the Datong station
图 4. 大通站旬尺度计算月径流模拟结果

Table 4. The monthly runoff simulation results of artificial neural network model calculated by ten-day scale [10]
表 4. 人工神经网络模型旬尺度计算月径流模拟结果[10]

站点	评价指标	屏山	宜昌	汉口	大通	平均
率定期	ART (%)	1.25	2.18	2.18	1.22	1.71
	AP (%)	100.00	97.50	97.50	100.00	98.75
检验期	ART (%)	15.92	21.56	21.56	14.72	18.44
	AP (%)	73.48	51.39	51.39	73.61	62.47

个站点的整个流量序列上，模型在中低流量区取得了较好的预测效果，在峰值预测上总体偏低，但总体取得了较好的预测效果。

总体而言，本文依据逐步回归方法支持向量机建立的耦合模型可以较好地用于长江流域中长期月、旬径流预报。同时，也为支持向量机这一机器学习方法用于其它流域的径流预测提供了一定的基础。

5. 结论

本文将支持向量机模型应用到长江流域中长期预报研究中，对长江流域屏山、宜昌、汉口、大通四个站点分别进行了月、旬径流预报，并且把月尺度模拟结果与旬尺度计算月径流结果进行比较，结果表明支持向量机模型模拟效果较好，率定期的平均相对误差较小，合格率较好，检验期的平均相对误差及合格率与率定期相比有一定的差距，旬尺度计算月径流有较好的预报精度。本次研究表明，支持向量机模型预测在径流预报中可以有较好的应用价值，为长江流域水文预报工作提供一定的参考。当然，本文中的支持向量机模型还有一定的优化空间，模型参数需要进一步的优化，以及模型精度的进一步提高，这些都有待研究。

基金项目

国家自然科学基金(51339004、51279139)；中央高校基本科研业务费专项资金(2042014kf1012、2042014kf0033)。

参考文献 (References)

- [1] 李红波, 夏潮军, 王淑英. 中长期径流预报研究进展及发展趋势[J]. 人民黄河, 2012, 34(8): 36-40.
LI Hongpo, XIA Chaojun and WANG Shuying. Research progress and development trend of medium-long term runoff forecast. Yellow River, 2012, 34(8): 36-40. (in Chinese)
- [2] 张兰影, 庞博, 徐宗学. 基于支持向量机的石羊河流域径流模拟适用性评价[J]. 干旱区资源与环境, 2013, 27(7): 113-118.
ZHANG Lanying, PANG Bo and XU Zongxue. Assessment on the applicability of support vector machine-based models for runoff simulation in Shiyang river basin. Journal of Arid Land Resources and Environment, 2013, 27(7): 113-118. (in Chinese)
- [3] BOX, G. E. P., JENKINS, G. M. Time series analysis forecasting and control. San Francisco: Holden-Day, 1976.
- [4] 王伟, 雷晓辉. 石羊河流域中长期径流预报模型的应用[J]. 人民黄河, 2014, 36(1): 42-44.
WANG Wei, LEI Xiaohui. Research on medium and long term hydrological statistics prediction in runoff forecasting in the Shiyang River Basin. Yellow River, 2014, 36(1): 42-44. (in Chinese)
- [5] 李继伟, 纪昌明. 基于支持向量机的水电站中长期径流组合预报[J]. 水电能源科学, 2013, 31(11): 13-16.
LI Jiwei, JI Changming. Medium and long-term runoff combination forecast based on support vector machine. Water Resources and Power, 2013, 31(11): 13-16. (in Chinese)
- [6] 崔东文, 郭荣. 基于几种参数优化的支持向量机在径流预报中的比较分析[J]. 水资源研究, 2013, 34(2): 34-38.
CUI Dongwen, GUO Rong. Supportvector machine (SVM) in the comparative analysis of runoff forecast based on several parameters optimization. Journal of Water Resources Research, 2013, 34(2): 34-38. (in Chinese)
- [7] 郭俊, 周建中. 基于改进支持向量机回归的日径流预测模型[J]. 水力发电, 2013, 36(3): 12-15.
GUO Jun, ZHOU Jianzhong. Daily runoff forecast based on improved support vector machine regression model. Water Power, 2013, 36(3): 12-15. (in Chinese)
- [8] 王双银, 杨筱筱. 基于相关系数和 Fisher 最优分割法的汛期分期研究[J]. 水文, 2011, 32(4): 1-5.
WANG Shuangyin, YANG Xiaoxiao. Flood season division based on correlation coefficient and optimum partition method of fisher. Journal of China Hydrology, 2011, 32(4): 1-5. (in Chinese)
- [9] 李琪, 马建斌, 刘洪吉. 基于逐步回归的投影寻踪水文预报模型研究[J]. 水电能源科学, 2011, 29(2): 10-12.
LI Qi, MA Jianbin and LIU Hongji. Research on projection pursuit forecast model based on stepwise regression. Water Resources and Power, 2011, 29(2): 10-12. (in Chinese)
- [10] 宋星原, 张利平, 张艳军. 长江流域水资源预测预报方案编制研究工作[R]. 武汉: 长江水利委员会水文局, 2014.
SONG Xingyuan, ZHANG Liping and ZHANG Yanjun. Water resources prediction scheme establishment research work in the Yangtze River Basin. Wuhan: Water Resources Commission Hydrology Bureau of Yangtze River, 2014. (in Chinese)