

Application of SSA-LSSVM in Mid-Long Term Runoff Prediction

Huanhuan Ba^{1,2}, Shenglian Guo^{1,2}, Yixuan Zhong^{1,2}, Zhangjun Liu^{1,2}

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

²Hubei Provincial Collaborative Innovation Centre for Water Resources Security, Wuhan Hubei

Email: hhba@whu.edu.cn

Received: Sep. 17th, 2016; accepted: Oct. 7th, 2016; published: Oct. 10th, 2016

Copyright © 2016 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

To improve the accuracy of runoff prediction, Singular Spectrum Analysis (SSA) is applied to preprocess the original flow series and a new reconstructed series is obtained. The monthly inflow data of the Shuibuya Reservoir from 1951 to 2009 were selected as a case study. Seasonal Autoregressive (SAR) model, support vector machine (SVM) and least square support vector machine (LSSVM) are used to simulate and predict the original and reconstructed data series. The results show that SSA-LSSVM performs the best among these models, in which the model efficiency coefficients reach 89% and 84% during the verification and testing periods, respectively. It is shown that the accuracy of mid-long term runoff prediction can be significantly improved by using SSA.

Keywords

Mid-Long Term Runoff Prediction, Singular Spectrum Analysis, SAR Model, SVM, LSSVM, Shuibuya Reservoir

SSA-LSSVM在中长期径流预测中的应用研究

巴欢欢^{1,2}, 郭生练^{1,2}, 钟逸轩^{1,2}, 刘章君^{1,2}

¹武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉

²水资源安全保障湖北省协同创新中心, 湖北 武汉

Email: hhba@whu.edu.cn

收稿日期: 2016年9月17日; 录用日期: 2016年10月7日; 发布日期: 2016年10月10日

作者简介: 巴欢欢(1992-), 女, 河南灵宝人, 硕士研究生, 主要从事水文预报方面研究。

摘要

为提高中长期径流预测精度,利用奇异谱分析(SSA)对输入资料进行数据预处理,消除噪声,得到重建序列。以水布垭水库1951~2009年的入库月径流资料为依据,选用季节性一阶自回归模型、支持向量机模型和最小二乘支持向量机模型作为径流预测模型,对原始序列和重建序列进行模拟预测。结果表明,基于奇异谱分析的最小二乘支持向量机的模拟预测精度最高,率定期和检验期的模型效率系数分别高达89%和84%。说明采用SSA对资料进行预处理可以显著提高中长期径流预报的精度。

关键词

中长期径流预测, 奇异谱分析, 季节性一阶自回归, 支持向量机, 最小二乘支持向量机, 水布垭水库

1. 引言

准确可靠的中长期径流预报对防汛抗旱、水库运行调度以及水资源规划管理等有着十分重要的意义[1]。中长期径流预报从传统的多元回归分析、逐步回归分析、自回归模型等模型方法到近年的人工神经网络、遗传规划、支持向量机等智能方法,涌现了很多研究成果[1]-[3]。其中支持向量机(SVM)是通过核函数实现高维空间的非线性映射[4],能够较好地描述径流的非线性特征,是水文领域近年来的研究热点。最小二乘支持向量机(LSSVM)采用二次规划方法将传统支持向量机中的不等式约束变为等式约束,提高了收敛速度,具有较好的非线性拟合能力[5]。然而径流时间序列本身具有典型的非线性和非平稳性特征,SVM和LSSVM模型虽然能够较好地描述径流的非线性特征,但在处理径流的非平稳性上存在缺陷。径流时间序列是被噪声污染的一系列准周期信号的组合[6],对输入序列数据进行预处理可以提高模型的预测精度。奇异谱分析(SSA),是一种有效的数据预处理技术,广泛应用于径流预报中[6]-[9]。因此本文运用SSA对原始序列数据进行预处理,以消除噪声,采用SAR(1)模型、SVM模型和LSSVM模型对水布垭水库1951~2009年的月入库径流原始序列和重建序列进行预测,将数据处理前后的三个模型预测结果进行比较,对比分析数据预处理对模型预测精度的影响。

2. 方法和模型

2.1. 奇异谱分析

奇异谱分析是一种广义的功率谱分析,具有稳定的识别和强化信号功能,其优点在于能够提取蕴含在时间序列中的不规则波动和随机性特征,消除噪声[10]。

2.1.1. 建立相空间

SSA的分析对象是一维时间序列,首先将所观测到的一维时间序列 x_1, x_2, \dots, x_N 按照给定的嵌套空间维数(即窗口长度)建立相空间轨迹矩阵:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{i+1} & \cdots & x_{N-L+1} \\ x_2 & x_3 & \cdots & x_{i+2} & \cdots & x_{N-L+2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_{i+L} & \cdots & x_N \end{bmatrix} \quad (1)$$

式中: \mathbf{X} 称为相空间中的轨迹矩阵; $L(1 < L < N)$ 称为窗口长度。

2.1.2. 奇异值分解

类似于常规的经验正交函数(Empirical Orthogonal Function, EOF)分解,轨迹矩阵 \mathbf{X} 也可以进行奇异值分解,

令 $S = \mathbf{X}\mathbf{X}^T$ ，可计算出矩阵 S 的特征值和其所对应的正交特征向量，则矩阵 X 的分量式为：

$$x_{i+j} = \sum_{k=1}^L a_i^k E_j^k \quad (2)$$

式中： $i=1, \dots, N-L+1; j=1, \dots, L; k=1, \dots, L$ ； E_j^k 是时间 EOF，记为 T -EOF， a_i^k 是时间主分量，记为 T -PC， a_i^k 是 E_j^k 表示的时间型在原序列 $x_{i+1}, x_{i+2}, \dots, x_{i+L}$ 时段的权重[11]。

2.1.3. 重建成分

SSA最重要的应用功能是通过重建成分(Reconstructed Components, RC)实现的，它在预报中用于提取有用信息，过滤掉噪声[11]。

根据最小二乘法建立最优规则，求解重建成分RC，由第 K 个 T -EOF 和 T -PC 重建 x_i 的成分记为 x_i^k ，将第 k 个重建成分用 x^k 表示，可以产生 L 个长度为 N 的时间序列 x^k ，即原始序列被分解为 L 个子序列的和，如下式所示：

$$x_i = \sum_{k=1}^L x_i^k \quad (3)$$

式中： $i=1, \dots, N$ 。

从序列中选取 $p (1 \leq p \leq L)$ 个具有贡献作用的有用成分进行重建，则 p 个序列之和等于重建序列，如下式所示：

$$x_i = \sum_{k=1}^p x_i^k \quad (4)$$

该重建序列是被 SSA 滤去噪声后的有用序列，蕴含原序列的周期成分。

2.2. 季节性一阶自回归模型

月径流由于有以年为周期的变化特点，其序列的统计参数与时间有关，故序列是非平稳的，考虑月径流的这种特性，采用广泛应用的月径流模型，季节性一阶自回归模型，SAR(1)，设 $X_{t,\tau}$ 为月径流序列， t 表示年份， τ 表示月份，则模型表达式为：

$$X_{t,\tau} = \varphi_{0,\tau} + \varphi_{1,\tau} X_{t,\tau-1} + \varepsilon_{t,\tau} \quad (5)$$

式中： $X_{t,\tau}$ 是 t 年 τ 月的径流量； $\varphi_{0,\tau}$ 和 $\varphi_{1,\tau}$ 为第 τ 月的两个回归参数； $\varepsilon_{t,\tau}$ 是 t 年 τ 月的独立随机项。参数 $\varphi_{0,\tau}$ 和 $\varphi_{1,\tau}$ 的计算公式为：

$$\varphi_{0,\tau} = \mu_\tau - \varphi_{1,\tau} \mu_{\tau-1} \quad (6)$$

$$\varphi_{1,\tau} = r_{1,\tau} \sigma_\tau / \sigma_{\tau-1} \quad (7)$$

式中： μ_τ 和 $\mu_{\tau-1}$ 分别为第 τ 、 $\tau-1$ 月份的径流量均值； σ_τ 、 $\sigma_{\tau-1}$ 为第 τ 、 $\tau-1$ 月份的径流量标准差； $r_{1,\tau}$ 为第 τ 月的一阶相关系数

$$r_{1,\tau} = \frac{\sum_{t=1}^n (X_{t,\tau} - \mu_\tau)(X_{t,\tau-1} - \mu_{\tau-1})}{(n-1)\sigma_\tau \sigma_{\tau-1}} \quad (8)$$

将所求的相关系数及参数代入式(5)中，即可得到预报方程，实现月径流预报。

2.3. 支持向量机模型

给定长度为 N 的训练集 $\{x_k, y_k\}_1^N$ ，其中 x 为 m 维输入向量， y_k 是 x_k 对应的输出值，支持向量回归的基本思想就是通过一个非线性映射 ϕ 将数据 x_k 映射到高维特征空间 F ，并在该空间进行线性回归，即

$$f(x) = w^T \cdot \phi(x) + b \quad (9)$$

式中: w 为超平面的权值向量, b 为偏置项。

根据支持向量机理论, 可将线性回归方程转化为优化问题求解, 目标函数如下:

$$\min_{w, b, \xi, \xi^*} J = \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (10)$$

约束条件:

$$\text{s.t.} \begin{cases} y_i - w^T \cdot \phi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*, i = 1, \dots, l \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (11)$$

式中: ξ , ξ^* 为松弛变量, ε 为 Vapnik- ε 不敏感代价函数所定义的误差, 常数 $C > 0$, 为惩罚函数。

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:

$$\begin{aligned} L = & \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l a_i (\xi_i + \varepsilon - y_i + w \cdot x_i + b) \\ & - \sum_{i=1}^l a_i^* (\xi_i^* + \varepsilon + y_i - w \cdot x_i + b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (12)$$

式中: a_i , a_i^* , η_i , η_i^* 都为拉格朗日乘子, 同时约束条件可转化为:

$$\text{s.t.} \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i^* \leq C \\ w = \sum_{i=1}^l (a_i - a_i^*) \phi(x_i) \end{cases}, i = 1, \dots, l \quad (13)$$

因此, 非线性回归问题可以通过解式(10)的对偶问题求解:

$$\max_{a, a^*} W(a_i, a_i^*) = -\frac{1}{2} \sum_{i, j=1}^l (a_i^* - a_i)(a_j^* - a_j) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (a_i + a_i^*) + \sum_{i=1}^l y_i (a_i - a_i^*) \quad (14)$$

求解此式即可得到回归函数:

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) + b \quad (15)$$

其中核函数 $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$ 是高维特征空间的内积, 满足 Mercer 条件。目前研究最多、最常用的核函数为高斯径向基函数, 本文亦采用此函数:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (16)$$

2.4. 最小二乘支持向量机

最小二乘支持向量机于 1999 年由 Suykens 等[12]提出, 它不仅具有支持向量机泛化能力强、全局最优等优点, 并且将支持向量机中求解二次规划问题转化为求解线性方程组问题, 因而极大地简化了计算, 提高了收敛速度[4]。其与支持向量机的不同点是在优化目标中选取了不同的损失函数, 最小二乘支持向量机的损失函数选取误差 ξ_i (允许错分的松弛变量) 的二范数。则最小二乘支持向量机根据结构风险最小化准则的优化问题为:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \frac{1}{2} \sum_{i=1}^l \xi_i^2 \quad (17)$$

约束条件为:

$$y_i - w^T \cdot \phi(x_i) = b + \xi_i, i = 1, \dots, l \quad (18)$$

通过引入拉格朗日函数来求解式(17) (18)的优化问题, 最优的拉格朗日乘子和 b 可根据 KKT(Karush-Kuhn-Tucker)条件和训练样本集 $\{(x_i, y_i)\}$ 求得, 核函数亦选取高斯径向基函数, 具体公式见文献[4], 即可得最小二乘支持向量机的回归函数模型。

LSSVM 用等式约束代替不等式约束, 将支持向量机中求解二次规划问题转化为求解线性方程组问题, 降低了计算的复杂性, 提高了收敛速度。

3. 实例分析

水布垭水库坝址位于清江流域中游的巴东县水布垭镇, 控制面积为 10,860 km², 水库正常蓄水位 400 m, 总库容 45.8 亿 m³, 多年平均流量 299 m³/s, 多年平均径流量为 94.4 亿 m³。采用水布垭水库 1951~2009 年(共计 59 年, 708 个月)的入库月径流资料进行分析研究。

3.1. 数据预处理

3.1.1. 数据标准化

为了消除数值不同量级之间的相互影响, 加快模型的识别精度与收敛速度, 对资料数据进行标准化处理[13]。

$$y_t = \frac{x_t - \bar{x}}{\sigma} \quad (19)$$

式中: x_t, y_t 分别表示 t 时刻的原始值和标准化后的月径流量, \bar{x} 表示月径流量的均值, σ 表示月径流量的标准差。

3.1.2. SSA 参数选择

运用 SSA 处理数据时, 窗口长度 L 和贡献成分个数 P 是两个重要的参数。窗口长度的选择直接影响着时间序列的有效分解, 决定能否准确捕捉时间序列的结构特征。窗口长度的选取没有一个统一的标准, 本文基于检验期不同窗口长度下模型的预测值与实测值的均方误差最小的原则[8]来选取窗口长度。考虑到月径流序列存在年周期, 具有季节性, 故本文在[3,12]之间选取最优窗口长度 L 。均方误差(Root Mean Square Error, RMSE)的计算式为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Q_{pt} - Q_{ot})^2} \quad (20)$$

式中: n 是检验期径流量个数, Q_{pt} 是检验期第 t 个时段径流预测值, Q_{ot} 是检验期第 t 个时段径流实测值。表 1 列出了检验期内三个模型在窗口长度 L 取[3,12]时的均方误差, 结果表明 SAR(1)、SVM、LS-SVM 三个模型分别在 $L = 3$ 、 $L = 11$ 、 $L = 11$ 时均方误差最小, 因此运用奇异谱分析对数据进行预处理时, SAR(1)模型、SVM 模型和 LS-SVM 模型的最优窗口长度分别取 3、11 和 11。

SSA 另一个重要的参数是贡献成分的个数 P 。给定一个窗口长度 L 后, 原始序列被分解为 L 个子序列, 从 L 个子序列中准确识别贡献成分是提取序列中的有用成分滤去噪声的关键。本文采用互相关函数法[9]来确定贡献成分 P 的个数。分别计算各个窗口长度下 L 个子序列与原始序列的互相关系数, 并将统计结果列于表 2。互相关系数为正的序列即为原始序列的贡献成分, 而互相关系数为负的序列即为噪声。如从表中知, 当 $L = 3$ 时, 子序列 2 和子序列 3 为贡献成分, 则将两者叠加构成重建序列 RC, 作为模型的输入。

Table 1. RMSE values of models under various window lengths L during testing periods**表 1.** 检验期内模型在各个窗口长度下均方误差

模型 \ L	3	4	5	6	7	8	9	10	11	12
SSA-SAR(1)	173.3	174.8	174.0	174.8	174.4	174.8	174.7	174.9	174.8	174.9
SSA-SVM	106.6	92.9	100.9	97.7	95.0	92.1	92.9	93.7	91.9	93.0
SSA-LSSVM	90.5	91.6	95.0	101.2	91.8	100.1	87.4	97.5	86.6	91.6

Table 2. Cross-correlation function values between subseries and original series under various window lengths L**表 2.** 各个窗口长度下子序列与原序列的互相关系数

L	序列编号												P
	1	2	3	4	5	6	7	8	9	10	11	12	
2	-0.20	0.68	—	—	—	—	—	—	—	—	—	—	1
3	-0.32	0.23	0.71	—	—	—	—	—	—	—	—	—	2
4	-0.33	-0.14	0.49	0.71	—	—	—	—	—	—	—	—	2
5	-0.32	-0.20	0.02	0.61	0.70	—	—	—	—	—	—	—	3
6	-0.31	-0.25	-0.07	0.14	0.65	0.70	—	—	—	—	—	—	3
7	-0.30	-0.26	-0.15	0.03	0.24	0.70	0.66	—	—	—	—	—	4
8	-0.27	-0.30	-0.20	-0.04	0.09	0.33	0.70	0.66	—	—	—	—	4
9	-0.26	-0.30	-0.23	-0.12	0.03	0.14	0.36	0.70	0.66	—	—	—	5
10	-0.26	-0.31	-0.26	-0.14	-0.05	0.08	0.17	0.34	0.68	0.66	—	—	5
11	-0.26	-0.30	-0.26	-0.16	-0.09	0.03	0.11	0.19	0.33	0.65	0.67	—	6
12	-0.26	-0.23	-0.22	-0.17	-0.15	-0.03	0.06	0.13	0.24	0.32	0.64	0.66	6

3.1.3. 预报因子的选取

中长期径流预报中，假设未来径流值与过去径流值之间存在某种确定的函数关系，利用过去的观测值估计未来值[14]，其数学描述为：

$$Q(t) = f(Q(t-1), Q(t-2), \dots, Q(t-m)) \quad (21)$$

式中： $Q(t)$ 为当前时段月径流量， $Q(t-m)$ 为当前时段径流的 m 个时段前期影响月径流量。

SAR(1)模型是用前一个月的径流预测当前时段的月径流，故预测因子为前一个月的径流。而SVM模型和LSSVM模型的预报因子的个数对训练结果有很大的影响，个数过小，则没有足够的序列信息让模型去捕捉，若个数过大，则会引入过多的噪声，不仅增加模型的训练时长，而且可能会干扰训练结果，因此采用SVM和LSSVM建模时，选择合适的时滞参数 m 至关重要[15]。本文采用自相关函数法[8]确定时滞参数 m ，图1和图2表示了原始径流序列与 $L = 11$ 时重建序列在滞时为1~24月的自相关函数值。从图1知，原始序列的自相关函数在 $m = 12$ 时取最大值，而图2中重建序列在 $m = 1$ 时自相关函数最大， $m = 12$ 时次之，然而预报因子越多，所包含的信息越多，且为方便比较，均选取前12个月的径流数据作为SVM和LSSVM模型原始序列和重组序列的预报因子。

3.1.4. 模型的评价指标

基于水文情报预报规范，本文选取以下四个指标作为预测模型的精度评价指标：

(1) 纳什模型效率系数

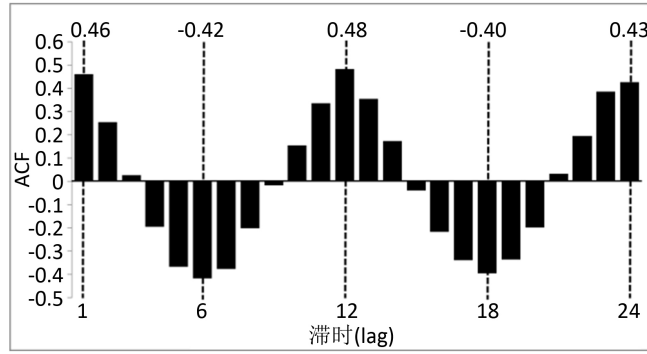


Figure 1. Autocorrelation function of the original series

图 1. 原始序列的自相关函数值

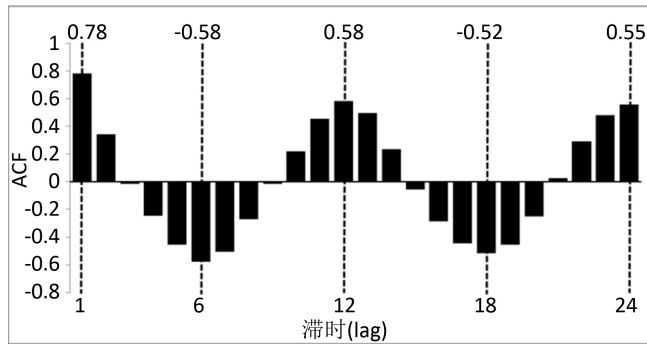


Figure 2. Autocorrelation function of the reconstructed series under $L = 11$

图 2. $L = 11$ 重建序列的自相关函数值

$$NS = \left(1 - \frac{\sum_{t=1}^n (Q_{ot} - Q_{pt})^2}{\sum_{t=1}^n (Q_{ot} - \bar{Q}_o)^2} \right)$$

(2) 水量平衡系数

$$WB = \frac{\sum_{t=1}^n Q_{pt}}{\sum_{t=1}^n Q_{ot}}$$

(3) 年均最大径流的相对误差

$$RE_{\max} = \frac{1}{l} \sum_{j=1}^l \left| 1 - \frac{Q_{pj,\max}}{Q_{oj,\max}} \right| \times 100\%$$

(4) 年均最小径流的相对误差

$$RE_{\min} = \frac{1}{l} \sum_{j=1}^l \left| 1 - \frac{Q_{pj,\min}}{Q_{oj,\min}} \right| \times 100\%$$

式中： n 是径流序列的长度， l 是年份数； Q_{pt} 和 Q_{ot} 分别表示 t 时刻径流的预测值和实测值； \bar{Q}_o 是实测径流均值； $Q_{pj,\max}$ 和 $Q_{oj,\max}$ 分别表示第 j 年最大径流的预测值和观测值； $Q_{pj,\min}$ 和 $Q_{oj,\min}$ 分别表示第 j 年最小径流的预测值和观测值。 NS 和 WB 越接近于 1，表示模型效果越好； RE_{\max} 和 RE_{\min} 越接近于 0，表示模型对极大值和极小值模拟预测能力越好。

3.2. 结果分析

本文选取水布垭水库 1951~2009 年的入库月径流资料进行分析, 将 1951~1998 年的月径流资料作为训练数据, 1999~2009 年的月径流资料作为检验数据。运用三个模型分别模拟预测原始序列和重建序列, 率定期和检验期的模拟预测结果见表 3, 并在图 3~5 中绘制了三个模型检验期 132 个月的预测值与实测值, 直观分析模拟预测效果。

3.2.1. SAR(1)模型模拟预测结果

如表 3 所示, 经过奇异谱分析预处理后, SAR(1)模型在训练期的模型效率系数由原来的 48% 提升到 56%, 水量平衡系数为 100%, 且年均最大径流相对误差和年均最小径流相对误差均有所降低, SAR(1)模型在训练期模拟精度虽有一定提升, 但模型的检验期在预处理前后, 预测精度变化不大, 这说明奇异谱分析对 SAR(1)模型模拟预测精度的提高影响不大。

3.2.2. SVM 模型模拟预测结果

由表 3 得出, 数据预处理前, SVM 模型率定期和检验期的纳什效率系数为 45%、43%, 检验期的年均最大和最小径流相对误差分别高达 29% 和 39%, 模型对洪水和枯水期的预测能力不足。经过奇异谱分析对原始数据进行预处理后, 模型的精度大幅度提升。SVM 模型检验期的纳什效率系数达 82%, 水量平衡系数也接近于 1;

Table 3. Performance results of models during training and testing periods

表 3. 率定期和检验期各模型的模拟结果

模型	率定期(1951.1~1998.12)				检验期(1999.1~2009.12)			
	NS	WB	RE _{max}	RE _{min}	NS	WB	RE _{max}	RE _{min}
SAR	0.48	1.00	0.29	0.43	0.36	1.18	0.42	0.34
SSA-SAR	0.56	1.00	0.21	0.34	0.36	1.18	0.40	0.29
SVM	0.45	0.85	0.33	0.34	0.43	1.00	0.29	0.39
SSA-SVM	0.83	0.99	0.13	0.36	0.82	1.03	0.12	0.42
LSSVM	0.43	1.00	0.32	0.82	0.38	1.16	0.31	0.64
SSA-LSSVM	0.89	1.00	0.11	0.47	0.84	1.04	0.09	0.59

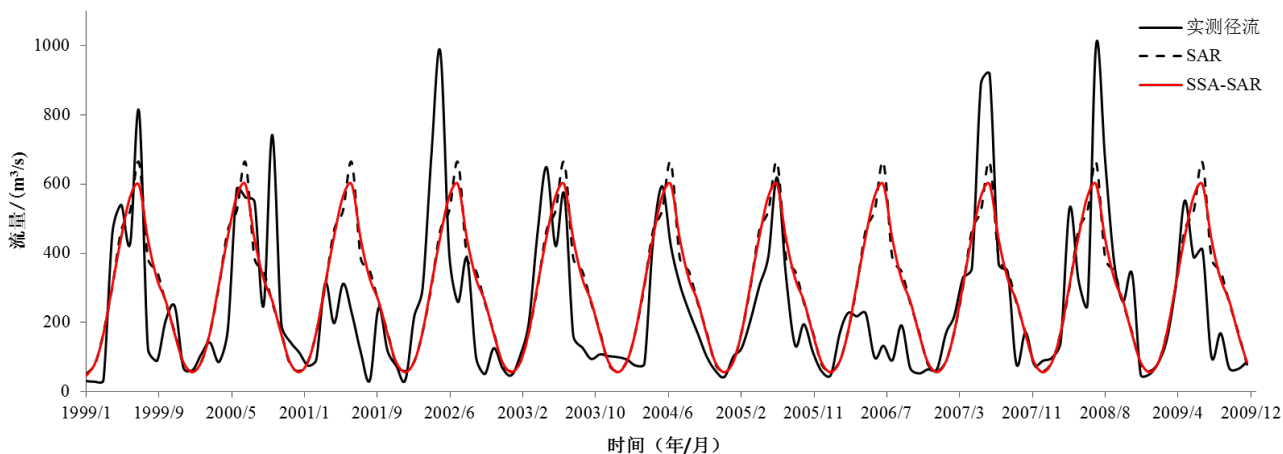


Figure 3. Comparison of observed inflow and predicted inflow by the SAR and SSA-SAR during testing period

图 3. 检验期模型 SAR 和 SSA-SAR 的预测值与实测值对比图

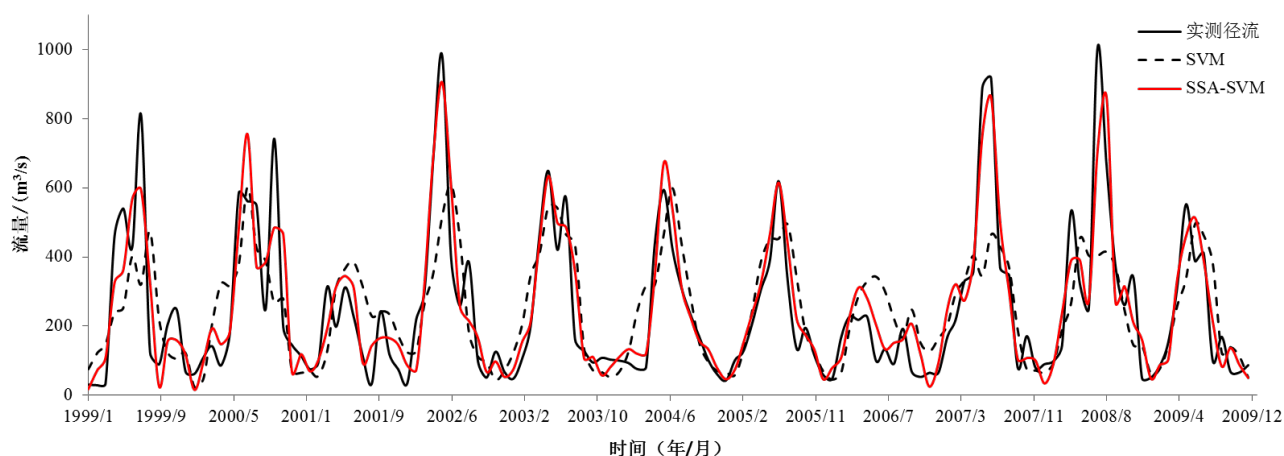


Figure 4. Comparison of observed inflow and predicted inflow by the SVM and SSA-SVM during testing period

图 4. 检验期模型 SVM 和 SSA-SVM 的预测值与实测值对比图

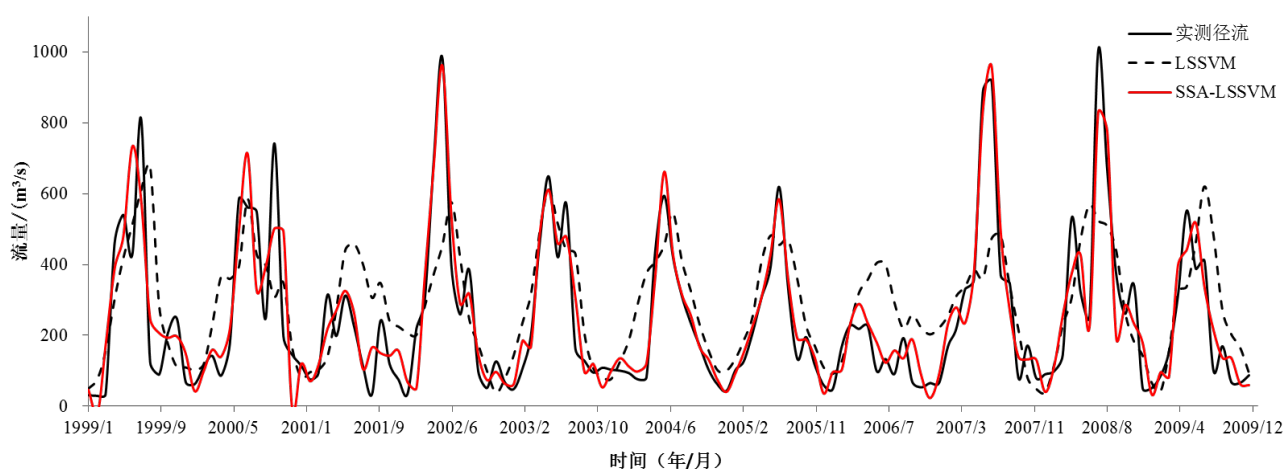


Figure 5. Comparison of observed inflow and predicted inflow by the LS-SVM and SSA-LSSVM during testing period

图 5. 检验期模型 LSSVM 和 SSA-LSSVM 的预测值与实测值对比图

检验期的年均最大径流相对误差从原有的 29%降低到 12%，然而年均最小径流相对误差相较于数据处理前略有增加。说明经过奇异谱分析后，SVM 模型的模拟预测精度显著提升，尤其对于年最大径流的预测，而对于枯水期的模拟预测未有明显改进。

3.2.3. LSSVM 模型模拟预测结果

由表 3 得出，数据预处理前，LSSVM 模型率定期和检验期的模型效率系数分别为 43%和 38%，检验期的年均最大和最小径流相对误差分别高达 31%和 64%，说明模型对洪水和枯水的预测能力不足。通过对原始输入数据序列采用 SSA 去除噪声后，将得到的重建序列作为模型的输入进行预测，模型在检验期的纳什效率系数达到 84%，水量平衡系数也较预处理前更接近于 1；检验期的年均最大径流相对误差从原有的 31%降低到 9%，预测精度显著提升，年均最小径流相对误差也从 64%降低到 59%，有所改进。从图 5 可以直观地看出，数据未处理前，LSSVM 模型预测值与实测值的总体趋势一致，但汛期的预测值与实测值偏差较大，尤其在 2002 年、2007 年和 2008 年三年的洪峰处最为明显，模型预测结果较差，而经过数据预处理后，模型的预测值与实测值拟合较好，对汛期的到达时刻和峰值模拟的较为准确，同时较好地模拟出了 2000 年、2002 年、2003 年、2008 年四年的双峰规律，说明经过奇异谱分析后，模型的预报精度显著提高。

4. 结论

本文以水布垭水库 1951~2009 年的月径流资料为依据, 采用奇异谱分析对其进行数据预处理, 选用 SAR(1) 模型、SVM 模型和 LSSVM 模型作为径流预测模型, 对原始序列和重建序列进行模拟预测。经过分析比较, 可以得出以下几条结论:

1) LSSVM 模型的模拟预测精度高于传统的 SAR(1)模型, 略低于 SVM 模型, 但 SVM 模型和 LSSVM 模型对月径流极大值和极小值的预测误差仍然偏大。

2) 奇异谱分析是一种有效的数据预处理技术, 经过预处理后的系列可以显著地提高 SVM 模型和 LSSVM 模型的模拟预测精度。

3) SSA-LSSVM 模型的模拟精度最好, 率定期和检验期的模型效率系数分别达到 89% 和 84%, 尤其表现对年最大月径流的预测, 显著地提高了水布垭水库中长期径流预报水平。

基金项目

国家自然科学基金重点项目(51539009)和十三五国家重点研发项目(2016YFC0402206)资助。

参考文献 (References)

- [1] 刘冀, 王本德, 袁晶瑄, 等. 基于相空间重构的支持向量机方法在径流中长期预报中应用[J]. 大连理工大学学报, 2008, 48(4): 591-595.
LIU Ji, WANG Bende, YUAN Jingxuan, et al. Application of support vector machine based on phase-space reconstruction to medium-term and long-term runoff forecast. Journal of Dalian University of Technology, 2008, 48(4): 591-595. (in Chinese)
- [2] 张卫国, 钟平安, 张玉兰, 等. 季节性支持向量机中长期径流预报模型[J]. 水力发电, 2014, 40(4): 17-21.
ZHANG Weigu, ZHONG Pingan, ZHANG Yulan, et al. A seasonal support vector machine mid-long term runoff forecast model. Water Power, 2014, 40(4): 17-21. (in Chinese)
- [3] 崔东文. 多隐层 BP 神经网络模型在径流预测中的应用[J]. 水文, 2013, 33(1): 68-73.
CUI Dongwen. Application of hidden multilayer BP neural network model in runoff prediction. Journal of China Hydrology, 2013, 33(1): 68-73. (in Chinese)
- [4] 周轶成, 焦国军. 基于最小二乘支持向量机的马营河中长期径流预测研究[J]. 甘肃水利水电技术, 2014, 50(12): 1-3.
ZHOU Yicheng, JIAO Guojun. Study on Mayinghe medium and long term runoff forecasting based on least square support vector machine. Gansu Water Resources and Hydropower Technology, 2014, 50(12): 1-3. (in Chinese)
- [5] 李佳, 王黎, 马光文, 等. LS-SVM 在径流预测中的应用[J]. 中国农村水利水电, 2008(5): 8-10.
LI Jia, WANG Li, MA Guangwen, et al. Application of least squares support vector machines in runoff forecast. China Rural Water and Hydropower, 2008(5): 8-10. (in Chinese)
- [6] WU, C. L., CHAU, K. W. Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. Journal of Hydrology, 2011, 399: 394-409. <http://dx.doi.org/10.1016/j.jhydrol.2011.01.017>
- [7] 汪芸, 郭生练, 李响. 奇异谱分析在中长期径流预测中的应用研究[J]. 人民长江, 2011, 42(9): 4-7.
WANG Yun, GUO Shenglian and LI Xiang. Application of singular spectrum analysis method in mid-long term runoff prediction. Yangtze River, 2011, 42(9): 4-7. (in Chinese)
- [8] WANG, Y., GUO, S., CHEN, H., et al. Comparative study of monthly inflow prediction methods for the Three Gorges Reservoir. Stochastic Environmental Research & Risk Assessment, 2014, 28(3): 555-570.
<http://dx.doi.org/10.1007/s00477-013-0772-4>
- [9] CHAU, K W, WU, C L. A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. Journal of Hydroinformatics, 2010, 12(4): 458-473. <http://dx.doi.org/10.2166/hydro.2010.032>
- [10] 丁裕国, 江志红. 气象时间序列信号处理[M]. 北京: 气象出版社, 1998.
DING Yugu, JIANG Zhihong. Signal processing on meteorological time series. Beijing: China Meteorological Press, 1998. (in Chinese)
- [11] 吴洪宝, 吴蕾. 气候变率诊断和预测方法[M]. 北京: 气象出版社, 2005.
WU Hongbao, WU Lei. Climate change rate of diagnosis and prediction method. Beijing: China Meteorological Press, 2005. (in Chinese)
- [12] SUYKENS, J. A. K., VANDEWALLE, J. Least squares support vector machine classifiers. Neural Processing Letters, 1999,

9(3): 293-300. <http://dx.doi.org/10.1023/A:1018628609742>

- [13] 邵骏, 袁鹏, 张文江, 等. 基于贝叶斯框架的 LS-SVM 中长期径流预报模型研究[J]. 水力发电学报, 2010, 29(5): 178-182. SHAO Jun, YUAN Peng, ZHANG Wenjiang, et al. Study of mid-long term runoff forecast based on LS-SVM in Bayesian evidence framework. Journal of Hydroelectric Engineering, 2010, 29(5): 178-182. (in Chinese)
- [14] 林剑艺, 程春田. 支持向量机在中长期径流预报中的应用[J]. 水利学报, 2006, 37(6): 681-686. LIN Jianyi, CHENG Chuntian. Application of support vector machine method to long-term runoff forecast. Journal of Hydraulic Engineering, 2006, 37(6): 681-686. (in Chinese)
- [15] 孙传文, 钟平安, 万新宇, 等. 考虑季节因子的支持向量机径流预测模型[J]. 中国农村水利水电, 2014(4): 101-104. SUN Chuanwen, ZHONG Pingan, WAN Xinyu, et al. Seasonal support vector machine model of the runoff forecast. China Rural Water and Hydropower, 2014(4): 101-104. (in Chinese)