DOI:10.11705/j.issn.1672-643X.2016.01.19

基于多重演化模式重构的水文预报模型及应用

王战平[1,2], 沈 冰[2], 吕继强[3]

(1. 西安浐灞生态区管理委员会, 陕西 西安 710024; 2. 西安理工大学 生态水利工程国家重点实验室培育基地, 陕西 西安 710048; 3. 长安大学 环境科学与工程学院, 陕西 西安 710064)

摘 要: 因长期受人类活动、气候变化等多重因素作用,水文时间序列表现出多时间尺度、多频率、动态变化、自记忆性等复杂性特征,增加了水文预报结果的不确定性。本文将经验模态分解模型,核主成分分析模型和支持向量机模型耦合,建立了针对复杂性水文时间序列的预报模型,并采用 NASH 效率系数、自相关系数、相对误差作为模拟预测精度及参数率定的多目标判断标准。模型应用于黄河花园口水文站径流序列的长期水文预报中,结果表明:模型预报时段长,具有较好的预测准确性和实践应用价值。该模型为多重因素作用的复杂性水文时间序列预报提供了一种方法。

关键词: 水文学及水资源; 水文预报模型; 模式重构

中图分类号:P338 **文献标识码**: A **文章编号**: 1672-643X(2016)01-0108-06

Application of hydrological forecast model based on multiple evolutionary reconstruction

WANG Zhanping[1,2], **SHEN Bing**[2], **LÜ Jiqiang**[3]

(1. *The Administration Committee of Xi'an ChanBa Ecological District*, *Xi'an* 710024, *China*;
2. *State Key Lab of Eco-Hydraulic Engineering*, *Xi'an University of Technology*, *Xi'an* 710048, *China*;
3. *School of Environmental Science and Engineering*, *Chang'an University*, *Xi'an* 710064, *China*)

Abstract: Due to the long-term effects of human activities and climate change, the hydrological time series exhibit more complicated variation features such as multiple time scales, many ruense dynamic change and self memory function, which increase the uncertainty of hydrological forecast result. The pape established a new hydrologic forecast model based on the empirical mode decomposition model, kernel principal component analysis model and support vector machine model, and choose Nash efficiency, self correlation coefficient, relative error as the multi objective criteria of forecasting precision and parameter calibration. The model was applied to the long-term runoff series at Huayuankou hydrology station of the Yellow River. The results show that the forecast time of the model is long and has better accuracy and practical value. The model can provide a method for the prediction of complex hydrological time series of multiple factors effect.

Key words: hydrology and water resources; hydrologic forecast model; evolution mode reconstruction

研究流域水文演化的趋势和方向,掌握水文序列变化的特性和规律,对科学合理的开展水文预报工作,特别是中长期水文预报,进而指导流域水资源管理和开发利用,促进社会经济发展,具有十分重要的意义。

由于流域水文演化受人类活动、气候变化等因素干扰,水文过程的长期演化规律具有明显的复杂性和不确定性特征。水文演化规律是对各种影响因子的综合作用结果,具有多时间尺度、多频率、动态变化、自记忆等特征。目前,涉及中长期水文时间序列预报的方法大多以水资源演变处于均衡稳定状态下为前提,以统计规律作为分析和计算的主要依据,

收稿日期:2015-09-20; **修回日期**:2015-12-19
基金项目:国家自然科学基金重点项目(50939004); 中央高校基本科研业务费专项资金资助项目(310829161006)
作者简介:王战平(1974-),男,西安临潼人,硕士,高级工程师,主要从事区域水环境与水资源研究工作。

往往造成研究或规划结果与实际状况脱离的问题。对于水文序列的非平稳性、外界因素与水文变化的影响机理较少考虑。例如:常用水文时间序列预测、预报的方法主要有 ARIMA、自记忆、灰色模型等[1-3]。气候变化与人类活动相互影响,尤其是人类活动干扰的影响,水文过程具有明显的非平稳特征,水文序列的复杂性和不确定性导致模型预测、预报的精度降低。近年来,考虑水文预报不确定性,较多采用智能优化方法提高水文预报精度,如采用人工神经网络、支持向量机等,其中应用混合智能预测模型,对各预测结果进行线性叠加,得到原始序列预测值的方法也得到较多尝试[4-12],但由于其对水文非平稳性、水文过程变化机理考虑不足,使得预测结果难以诠释,预报不确定性较大,实践中应用较少。

本文综合考虑水文变化的复杂性和不确定性,建立基于流域水文形成外界条件与水文变化之间映射关系的水文预报模型,以期提高水文预报准确性。模型应用结果对流域的水资源规划与管理、生态环境的可持续改善具有重要的参考价值。

1　基于多重演化模式重构的水文预报模型

基于多重演化模式重构的水文预报模型,是将经验模态分解模型(Empirical Mode Decomposition,简称 EMD),核主成分分析模型(Kernel principal component analysis,简称 KPCA)和支持向量机模型(Support Vector Machine,简称 SVM)耦合,建立针对复杂性水文时间序列的水文预报模型:即在应用 EMD 揭示复杂水文时间序列的多重演化模式基础上,将水文时间序列分解为不同演化模式序列,应用 KPCA 模型筛选和重构预报模式集合,作为 SVM 预测模型的多元输入,最后输出水文主序列的预报值。采用 NASH 效率系数、自相关系数、相对误差作为模拟预测精度的判断标准,进行模型参数率定。

1.1　经验模态分解

EMD 模型对实际信号进行时频分析时的假设条件为:信号由多个本征模态信号(Intrinsic Mode Signal,简称 IMS)或本征模态函数(Intrinsic Mode Function,简称 IMF)组成,如果本征模态信号相互重叠则形成复合信号[4]。应用 EMD 方法可以将水文时间序列分解为多个 IMF 分量和趋势项,IMF 分量代表不同特征尺度(频率)的数据序列,可以是线性的也可以是非线性的。对于时间序列 $x(t)$,EMD 分解具体步骤如下:

首先,采用三次样条函数拟合序列 $x(t)$ 的极大和极小值点,确定上、下包络线,并计算其均值 l,则移除低频的新数据序列 m 可表示为:

$$m = x(t) - l \tag{1}$$

对 m 重复以上过程 k 次,直至平均包络趋于零为止,所得数据记为 m_k。

$$m_k = m_{k-1} - l_k \tag{2}$$

式中:m_k、l_k 为第 k 次计算数据和均值;m_{k-1} 为第 $k-1$ 次计算数据。

设定计算限制标准差 D,判断计算结果是否为 IMF 分量。在满足 $m_{(k)}(n)$ 足够接近 IMF 同时,控制分解次数,使得 IMF 分量保持原始信号中幅值调制的信息。

$$D = \sum_{k=1}^{N} \frac{|\, m_{(k)}(n) - m_{(k-1)}(n) \,|^2}{m^2_{(k-1)}(n)} \tag{3}$$

式中:N 为序列长度。限差 D 的一般取值区间为 $[0.2, 0.3]$,当 m_k 满足限制条件时,可视为一个 IMF 分量,记为 F_1。其代表了原始序列的高频成分,从原始序列 $x(t)$ 中去除高频组分,得到新的差值序列。

$$r_1 = x(t) - F_1 \tag{4}$$

r_1 中包含较长的局部特征时间尺度信息,重复以上步骤,直至 r_1 中仅包含原始序列的趋势信息为止,依次得到 $x(t)$ 的本征模态函数分量 $F_1, F_2, \cdots, F_n$ 和趋势项 r_n。原始序列可表示为:

$$x(t) = \sum_{j=1}^{n} F_i + r_n \tag{5}$$

1.2　核主成分 - 支持向量机(KPCA - LSSVM)预测模型

将 EMD 函数分解结果中本征模态函数分量 $F_1, F_2, \cdots, F_n$ 和趋势项 r_n 作为输入变量,水文时间序列主序列值为预报模型输出,建立基于演化模式分解的核主成分 - 支持向量机(KPCA - LSSVM)预报模型。

1.2.1　*核主成分模型(KPCA)*　这里将 EMD 函数分解结果中本征模态函数分量 $F_1, F_2, \cdots, F_n$ 和趋势项 r_n 数据集简化为 $F_i, i = 1, \cdots, n, F_i \in R^N$,通过非线性映射 ϕ,建立输入样本到高维特征空间 H 的映射关系。计算映射后样本的协方差矩阵:

$$\theta = \frac{1}{n} \sum_{i=1}^{n} \phi(F_i) \phi(F_i)^{\mathrm{T}} \tag{6}$$

协方差矩阵的特征向量及特征值表示为 Ψ, λ。对 $\lambda \neq 0$ 时存在式:

$$\Psi = \frac{1}{n} \sum_{i=1}^{n} \theta \phi(F_i) \tag{7}$$

将式(7)左乘 $\phi(F_i)$ 可得:

$$\lambda\langle\phi(F_i),\Psi\rangle = \langle\phi(F_i),\lambda\Psi\rangle \tag{8}$$

引入核函数 Ω_{ij},即:

$$\Omega_{ij} = \Omega(F_i,F_j) = \langle(\phi(F_i),\phi(F_j)\rangle$$
$$i,j = 1,2,\cdots,n \tag{9}$$

式中:θ_i 为方程系数。求解式 $n\lambda\theta = \Omega\theta$,得到特征值及对应的特征向量。映射后的样本 $\phi(F_i)$ 在特征向量 Ψ 上的投影为:

$$\varphi_k(F) = \langle\Psi,\phi(F)\rangle = \sum_{i=1}^{n}\theta\phi(F_i)\phi(F)$$
$$= \sum_{i=1}^{n}\theta_i\Omega(F_i,F) \tag{10}$$

式中:$k = 1,2,\cdots,p$,称 $\varphi_k(F)$ 为对应于 $\phi(F)$ 的第 k 个非线性主成分分量。p 个特征向量的提取原则如下:

$$\eta = \sum_{k=1}^{p}\theta\lambda_k \Big/ \sum_{i=1}^{N}\lambda_k \geqslant \varepsilon \tag{11}$$

式中:η 为特征值的方差贡献率;ε 为选取的百分比值,本次应用取90%。当 $\eta \geqslant \varepsilon$ 时,即可将前 p 个特征值所对应的特征向量构成的空间作为低维投影空间,完成 p 个主成分提取。

1.2.2 支持向量机模型(SVM) 假设给定 n 个训练数据样本为:$\{(F_i,X_i),i = 1,2,\cdots,n,\ F_i \in R^N,\ X_i \in R^L\}$,在特征空间中最优决策函数为:

$$X_i = w^{\mathrm{T}}\phi(F_i) + a \tag{12}$$

式中:$\phi(\cdot)$ 是输入数据到高维特征空间 H 的非线性映射;w、a 分别为权矢量和函数偏置量。

函数逼近的最小二乘支持向量机优化问题可转化为:

$$\min L(w,\zeta) = \frac{1}{2}w^{\mathrm{T}}w + c\sum_{i=1}^{n}\zeta_k^2 \tag{13}$$

式中:c 为正规化参数;ξ 为松驰变量;高维空间优化中内积运算采用的核函数为高斯径问基函数 $\Omega(F,X)$。

高维空间优化中内积运算采用的核函数为高斯径向基函数 $\Omega(F,X)$。

SVM 模型回归函数为:

$$H(F) = \sum_{i=1}^{n}b_i\Omega(F,X) + a \tag{14}$$

式中:b_i 为拉朗日乘子。

1.2.3 模型参数率定 建模过程中,首先输入本征模态函数分量 $F_1,F_2,\cdots,F_n$ 和趋势项 r_n,经过核主成分分析剔除重复相关及噪声影响,在满足累计方差贡献率要求的同时,达到较好的数据降维效果,然后进行多元支持向量机模型拟合及预测。

最后,若当模拟精度满足要求时演化模式重构及模型参数确定,预报模型初步建立完成,否则重复核主成分-支持向量机的模型调参与模拟,直至模拟精度满足要求,模型建立完成。

1.2.4 模型预测精度评价标准 采用相对误差、相关系数和 NASH 效率系数进行结果评价计算。考虑水文序列非平稳性和阶段性趋势变化,单一指标评价结果很难反映和指导长时期的水文序列模拟。对于较长时期水文模拟结果评价,可以采用多种简单方法对比较为合理。

相对误差

$$R = \frac{|Q_r(i) - Q_s(i)|}{Q_s(i)} \times 100\% \tag{15}$$

式中:$Q_r(i)$ 为实测流量;$Q_s(i)$ 为模型模拟流量。R 越小则表明模拟与实测的径流总量越接近。

(2)NASH 效率

$$\mathrm{NASH} = 1 - \frac{\sum_{i=1}^{n}[Q_r(i) - Q_s(i)]^2}{\sum_{i=1}^{n}[Q_r(i) - \overline{Q}_0]^2} \tag{16}$$

式中:$\overline{Q}_0$ 为实测流量系列的平均值;NASH 效率系数表示径流预报与实测值之间的吻合程度。NASH 效率系数越接近 1 则模拟流量过程越接近观测,模拟与实测序列的相关性较好,同时模拟误差较小。

(3)相关系数

$$C = \frac{\sum_{i=1}^{n}(Q_r(i) - \overline{Q}_r)(Q_s(i) - \overline{Q}_s)}{\sqrt{\sum_{i=1}^{n}(Q_r(i) - \overline{Q}_r)^2}\sqrt{\sum_{i=1}^{n}(Q_s(i) - \overline{Q}_s)^2}} \tag{17}$$

式中:C 越小,表明模型模拟与观测序列的对应关系越好。

2 实例应用

将基于多重演化模式重构的中长期水文预报方法应用于黄河花园口水文站 1919-2004 年,共计 86 年天然径流量年值预报中。花园口水文站位于黄河中、下游的分界点附近,上距河源约 4 696 km,下距河口 767.7 km,集水面积 73 万 km^2,占黄河流域总面积的 97%,是黄河下游最重要的水沙控制站。近年来,由于黄河上、中流域气候变化和水资源开发利用强度提高等原因,黄河下游水量减少,河道过洪能力锐减。区域水文过程受人类活动干扰较

大,水文序列变化规律复杂,水文预报的不确定性强。利用经验模态分解(EMD)方法分析在多重因素长期影响下,年径流序列的多时间尺度演化模式变化情况,并基于分解序列,建立 KPCA – LSSVM 水文预报模型。同时,将本文提出的多输入单输出的 KPCA – LSSVM 预报模型与一般单输入单输出 SVM 预报模型、多输入多输出 KPCA – LSSVM 预报模型的预报效果,进行比较分析。

2.1　EMD 分解模式及分析

黄河花园口年径流量系列的 EMD 分解结果(图 1)可以看出,径流量序列内在演化包含 5 个本征模态函数(IMF1 ~ 5)和趋势项(Res)。各分量代表不同时间尺度的演化过程,有不同的变化频率,变化频率在长期的演化过程中有所改变。

趋势项说明径流量序列自 20 世纪 90 年代,已经较长时间处于缓慢下降过程中。IMF1 分量变化频率较高,可以认为是随机因素影响的结果。IMF2 ~ IMF5 具有较大时间尺度的变化过程,反映了年径流量与影响其长期演化的主要因子之间的响应关系。长期变化过程看,不同频段的演化模式均从 20 世纪 90 年代开始发生较大变化,年径流序列各分量的振动频率和振幅均出现明显减弱现象,主要变化周期呈现衰减过程。结合已有研究成果[11-12],认为径流序列变化与其影响因素作用相对应,由于黄河流域水资源开发利用强度的提高、流域水资源需求量逐年增加、产汇流下垫面条件改变以及近几十年的气候变化等因素的影响,已经使年径流序列的演化模式发生改变。

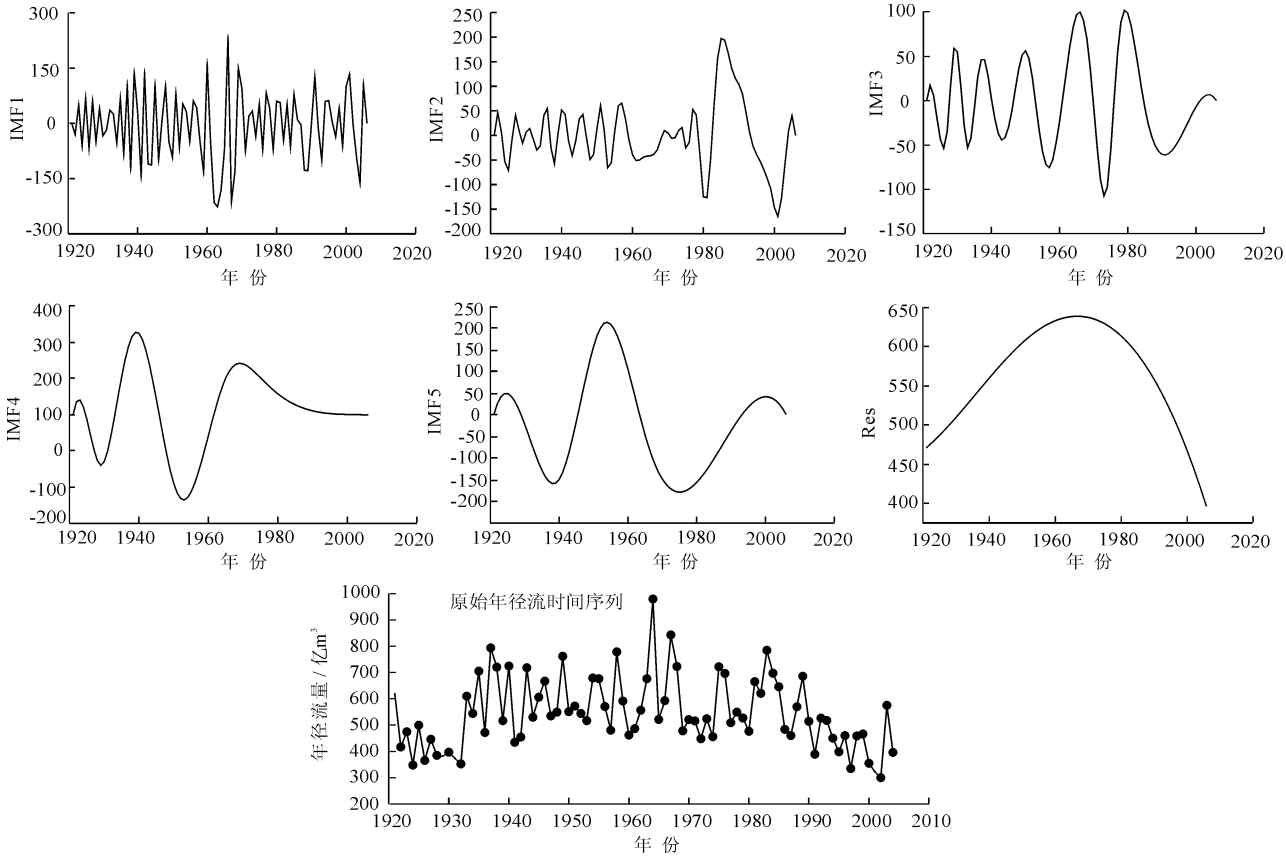

图 1　花园口水文站年径流演化模式分解

2.2　预报模型模拟

依据 KPCA 模型原理,选取累计方差贡献率大于 90% 的演化模式特征重构即 IMF2 ~ IMF5 及 Res,作为 SVM 模型输入项,建立多元非线性 KPCA – SVM 智能预报模型。按照已得到多元输入项与水文模拟序列的非线性映射关系,进行 30 年模拟和模型参数调节,最终获得预报结果。图 2、3 分别为预报模型在黄河花园口水文站年径流序列的模拟及预报效果。

表 1 为依据相对误差 R、相关系数 C 和 NASH 效率系数进行结果评定。依据《水文情报预报规范》(GB/T 22482 – 2008,水文情报预报规范)要求,径流预报以实测径流的误差 20% 作为许可精度,评价模型模拟与预测效果是否满足要求。

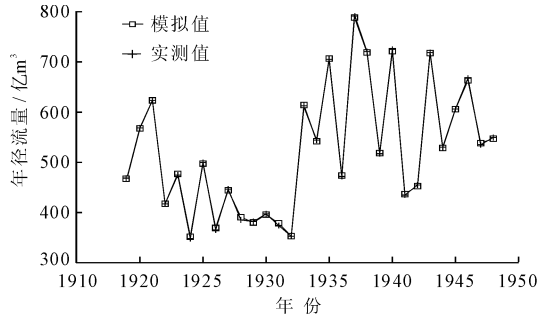

图 2 花园口水文站年径流模拟结果

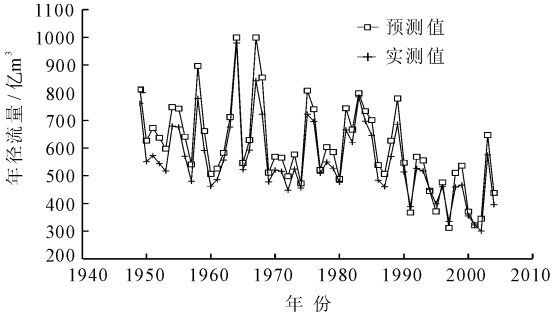

图 3 花园口水文站年径流模拟结果

表 1 中长期水文预报模型模拟及预测结果检验

时段划分	黄河花园口水文站以上流域	评价结果			
		C	*R*/%	NASH	是否满足20%要求
模拟期 30 年	1919 – 1948	0.99	0.46	0.99	是
预测期 56 年	1949 – 2004	0.03	8.32	0.99	是

2.3 模型预报结果对比

对基于支持向量机模型的不同预报方法进行建模研究,并对比应用结果的优劣。对比本文建立的基于演化模式重构的多元非线性 KPCA – LSSVM 预报模型(简称:非线性 EMDSVM)、一般单输入单输出 SVM 预报模型、基于演化模式重构的多元线性叠加预报模型(简称:线性叠加 EMDSVM)的预报效果。图 4 为不同模型的预测结果对比。表 2 为依据不同标准的精度评定结果。

对比表 2 和图 4 的预测结果,本文建立的非线性 EMDSVM 预报模型效果明显优于其他两种方法。水文时间序列变化受人类活动、气候变化影响的叠加作用,变化特性十分复杂,水文预报的不确定性大,影响了一般水文预报模型的预报效果。与其他方法相比,本文中采用的 KPCA 模型对本征模态函数分量 $F_1, F_2, \cdots, F_n$ 和趋势项 r_n 进行优选,经过核主成分分析剔除重复相关及噪声项 IMF1 分量的影响,实现演化模式优选和重构,然后进行模型参数优化和模拟计算,最终得到较好的预报精度。

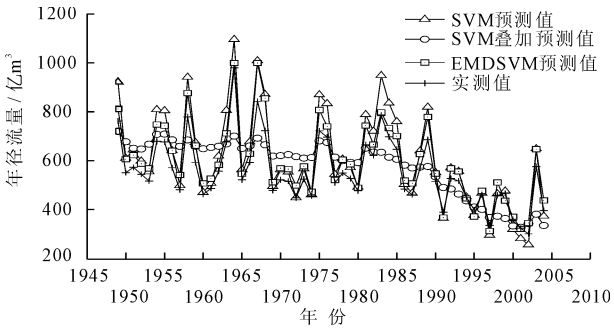

图 4 基于不同重构模式的花园口水文站年径流预测结果比较

表 2 中长期水文预报模型模拟及预测结果检验

预报模型类别	预测时段(56 年)	评价结果		
		C	*R*/%	*NASH*
非线性 EMDSVM	1949 – 2004	0.98	8.3	0.89
一般 SVM	1949 – 2004	0.99	10.6	0.79
线性叠加 EMDSVM	1949 – 2004	0.68	15.9	0.24

3 结果与讨论

本文针对现有长期水文预报模型存在的问题,建立经验模态分解模型与核主成分分析模型、支持向量机模型的耦合智能预报模型,应用于复杂性较强的水文过程预报中,得到以下主要结论:

(1)长期水文预报是流域水资源管理的重要依据。全面准确的把握未来水文情势变化,有助于合理科学的解决流域水资源紧缺、水资源开发利用不合理等问题。由于水文变化长期受到气候变化及人类活动等多重因素影响,水文系统内部发生变化,水文演化过程呈现复杂和多变特征,传统水文预报方法的预报效果难以满足社会经济发展的需求。因而,基于流域水文变化机理,采用智能优化方法,建立符合流域水文变化的非线性水文预报模型,是提高预报精度的有效途径之一。通过实例检验,本文建立的多元演化模式重构的 EMDSVM 预报模型体现出很好的非线性信息处理能力,对非线性的水文时间序列具有较好的拟合和预报能力。

(2)依据流域水文特性,分析多个不同频率尺度的水文演化分量、随机序列和趋势项的演化过程,及主要因子与各演化模式之间的响应关系,从整体和局部把握水文演化模式。为水文预报模型奠定良好的机理基础;根据各演化模式分量的变化特性,提取影响水文序列变化的主要模式分量,剔除水文演化中无贡献率的干扰因素。减少支持向量机模型预报过程中的盲目性,预报时段长度和预报精度有较大提高。

实例研究表明,考虑水文过程在不同时间尺度上序列变化的时频规律,建立可以反映水文变化机理的非线性水文预报模型,可提高模型模拟及预报精度,预报更加合理,且预报时长为模拟时段的 2 ~ 3 倍。可一定程度上减少过度模拟、预测预报精度降低较快等问题对水文预报结果的影响。

参考文献:

[1] 丁 晶,邓育仁. 随机水文学[M]. 成都:成都科技大学出版社,1988.

[2] 李荣峰,沈 冰,张金凯. 基于相空间重构的水文自记忆预测模型[J]. 水利学报,2006,37(5):583 - 587.

[3] 王文圣, 金菊良, 李跃清. 水文随机模拟进展[J]. 水科学进展, 2007,18(5): 768 - 775.

[4] 李 强,王太勇,王正英,等. 基于 EMD 和支持向量数据描述的故障智能诊断[J]. 中国机械工程,2008,19(22):2718 - 2721.

[5] Alemu E T, Palmer R N, Polebitski A, et al. Decision support system for optimizing reservoir operations using ensemble streamflow predictions [J]. Journal of Water Resources Planning and Management - Asce, 2011,137(1): 72 - 82.

[6] Sajikumar N, Thandaveswara B S. A non-linear rainfall-runoff model using an artificial network [J]. Journal of Hydrology, 1999,216(1):32 - 55.

[7] Pal I, Lall U, Robertson A W, et al. Predictability of western himalayan river flow: melt seasonal inflow into Bhakra Reservoir in Northern India [J]. Hydrology and Earth System Sciences Discussions, 2012,9(7):8137 - 8172.

[8] Mitani Y, Tsutsumoto K, Kagawa N. Time series prediction of acoustic signals using neural network model and wavelet shrinkage [C]//. Processings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden: IIAV, 2003:4189 - 4196.

[9] Boucher M A, Tremblay D, Delorme L, et al. Hydro-economic assessment of hydrological forecasting systems [J]. Journal of Hydrology, 2012,416 - 417:133 - 144.

[10] Liu Xingjie, Mi Zengqiang, Yang Qixun, et al. A novel multi - step prediction for wind speed based on EMD [J]. Transactions of China Electrotechnical Society, 2010,25(4):165 - 170.

[11] 曹鸿兴,郑小康,杨志峰,等. 黄河天然径流量变化趋势及其影响分析[J]. 北京师范大学学报(自然科学版), 2009,45(1):80 - 85.

[12] 郑红星,刘昌明. 黄河流域水资源演化模式分析[J]. 地理学报, 2004,59(2):267 - 273.

(上接第 107 页)

[11] 马海娇,严登华,翁白莎,等. 典型干旱指数在滦河流域的适用性评价[J]. 干旱区研究, 2013, 30(4):728 - 734.

[12] 温克刚. 中国气象灾害大典(新疆卷)[M]. 北京:气象出版社, 2006.

[13] 新疆维吾尔自治区统计局. 新疆统计年鉴[M]. 北京:中国统计出版社.

[14] McKee T B, Doesken N J, Kleist J. The relationship of drought frequency and duration to time scales [C]. //Proceedings of the 8th Conference on Applied Climatology. USA: American Meteorological Society, 1993.

[15] 王 莺,李耀辉,胡田田. 基于 SPI 指数的甘肃省河东地区干旱时空特征分析[J]. 中国沙漠,2014,34(1):244 - 253.

[16] 马明卫,宋松柏. 渭河流域干旱指标空间分布研究[J]. 干旱区研究, 2012,29(4):681 - 691.

[17] Vicente - Serrano S M, Beguería S, López - Moreno J I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index [J]. Journal of Climate, 2010, 23(7): 1696 - 1718.

[18] Vicente - Serrano S M, Beguería S, López - Moreno J I, et al. A new global 0.5° gridded dataset (1901 - 2006) of a multiscalar drought index: Comparison with current drought index datasets based on the Palmer Drought Severity Index [J]. Journal of Hydrometeorology,2010,11(4): 1033 - 1043.

[19] Thornthwaite C W. An approach toward a rational classification of climate [J]. Geographical Review, 1948, 38(1):55 - 94.

[20] 庄少伟,左洪超,任鹏程,等. 标准化降水蒸散指数在中国区域的应用[J]. 气候与环境研究,2013,18(5) : 617 - 625.

[21] 王 林,陈 文. 标准化降水蒸发指数在中国干旱监测的适用性分析[J]. 高原气象, 2014,33(2):423 - 431.

[22] 黎小燕,吴志勇,陆桂华. 三种干旱指数在西南地区的应用及相关性分析[J]. 水电能源科学, 2014, 32(5):1 - 5.

[23] 石 崇,刘晓东. 1947 - 2006 年东半球陆地干旱化特征——基于 SPEI 数据的分析[J]. 中国沙漠,2012,32(6):1691 - 1701.

[24] 赵 静,严登华,杨志勇,等. 标准化降水蒸发指数的改进与适用性评价[J]. 物理学报,2015,64(4):049202 - 1 - 049202 - 9.