文章编号：1000-8608(2012)01-0079-06

基于相关向量机的中长期径流预报模型研究

仕玉治[1,2]， 彭 勇*[1]， 周惠成[1]

（1.大连理工大学 水利工程学院，辽宁 大连 116024；
2.山东省水利科学研究院，山东 济南 250013）

摘要： 鉴于其优越的预报性能，将相关向量机(RVM)应用到中长期径流预报中，并在相空间重构的基础上，建立了基于相关向量机的径流预报模型.该模型首先对径流时间序列进行相空间重构，并以重构后的径流序列作为模型输入；其次，采用粒子群优化(PSO)算法识别模型参数，利用优化所得重构参数验证时间序列具有混沌特性，在模型内循环过程中采用EM算法迭代估计超参数，并将RVM与应用较为广泛的最小二乘支持向量机(LSSVM)和自动回归滑动平均模型(ARMA)进行了比较分析，结果表明该模型具有较好的泛化能力；最后，基于水文过程变化的不确定性、RVM描述输出值的不确定度以及相应概率下的预报区间，使得调度人员在决策中能考虑预报的不确定性，定量估计各种决策的风险和效益.

关键词： 相空间重构；相关向量机；长期径流预报；PSO算法

中图分类号： TV124 **文献标志码：** A

0 引 言

中长期径流预报对统筹安排防洪与抗旱、水库调度与管理等事务，实现水资源的最大利用效益具有十分重要的实践意义.由于水文系统本身的非线性及水文要素变化的不确定性，目前基于严密的物理方法还很难对径流等水文现象进行描述和预测，人们主要借助于成因分析法、水文统计法、模糊分析等方法来描述和预测水文过程.水文统计法依据水文资料的统计规律进行预测，方法较为常用，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，并建立模型进行预测，如时间序列分析法；另一类是分析水文要素与多因子之间的相关关系，建立模型进行预测，如多元回归法.两类均可直接利用原序列建立线性或非线性关系进行预测，但其精度有时不能满足工程需要.相空间重构成为分析时间序列的一种崭新的方法，通过挖掘或者恢复水文系统的多变量影响因子，重构水文非线性动力系统，国外许多学者对短期径流预报进行研究，并取得了较好的应用效果[1,2].时间滞时 τ 和嵌入维数 m(文中也称重构参数)对时间序列的噪声和数据量大小等影响因素比较敏感[3]，通常采用互信息法、关联维数法、虚假近邻法和Cao法等多种方法所得到的估计值差别较大，不利于获得较好的预报精度.本文采用Yu等[4]和Sivakumar[5]提出的方法，优化得到重构参数.

一般线性方法难以描述水文系统非线性特征，许多新的方法逐步被引用到水文预报模型当中，如贝叶斯理论、人工神经网络(ANN)、支持向量机(SVM)等[2,6,7]，进一步发展了非线性径流预报模型.2000年Tipping提出了一种新的稀疏概率模型相关向量机[8](relevance vector machine，RVM)，该方法用非线性核函数映射到高维空间，在高维空间进行线性回归，成功实现非线性向线性转化，同时基于贝叶斯理论定义模型参数，不仅可以定量预报，而且能够以概率分布的形式描述水文预报不确定度，可为水库调度决策分析提供更多的可利用信息.目前，其已应用到图像分析[9,10]、信道均衡[11]等分类与回归问题，获得了较好的应用效果.

综上所述，确定自身前期影响因子和建立预报模型，是时间序列分析预测方法的关键，本文首先对径流时间序列进行相空间重构，挖掘水文系统多变量因子；然后利用重构后的时间序列建立

收稿日期： 2010-01-20； **修回日期：** 2011-10-11.
基金项目： 水利部公益性行业专项资助项目(201001024)；国家自然科学基金资助项目(51109025).
作者简介： 仕玉治(1980-)，男，博士，工程师，E-mail：syz101066@163.com；彭 勇*(1979-)，男，博士，E-mail：pyongcuidi@163.com.

RVM 非线性径流预报模型，并采用粒子群优化(PSO)算法[12]辨识模型参数；最后应用实例验证本文模型的有效性.

1 混沌时间序列相空间重构

相空间重构是混沌理论的基础，依据 Takens 理论[13]，对某一混沌时间序列$\{x_i: i=1,2,\cdots,n\}$，只要适当选取时间滞时 τ 和嵌入维数 m，且嵌入维数满足 $m \geqslant 2D+1$，其中 D 为饱和关联维数，即可重构与原未知动力系统具有相同几何特征的 m 维相空间，则相空间中的点可以表示为 $\boldsymbol{X}_i = (x_i \quad x_{i+\tau} \quad x_{i+2\tau} \quad \cdots \quad x_{i+(m-1)\tau})^{\mathrm{T}}$，由 N 个相点组成的延迟状态向量表示为$\{\boldsymbol{X}_i: i=1,2,\cdots,N\}$，其中 $N=n-(m-1)\tau$，则相应关联积分表达式为

$$C(r,m)=\frac{2}{(N+1)N}\sum_{i=1}^{N}\sum_{j=i+1}^{N}H(r-\|\boldsymbol{X}_i-\boldsymbol{X}_j\|) \tag{1}$$

式中：$H(x)=\begin{cases}0; & x\leqslant 0\\ 1; & x>0\end{cases}$，$r$ 为标度尺度，$\|\cdot\|$为欧几里得范数.

对于混沌时间序列，关联积分 $C(r,m)$ 与标度尺度 r 近似成指数关系：$C(r,m)\propto r^D$，

$$D=\lim_{r\to 0}\lim_{m\to\infty}\frac{\partial \ln C(r,m)}{\partial \ln r} \tag{2}$$

2 相关向量机(RVM)径流预报模型

已知相关向量$\{\boldsymbol{X}_i: i=1,2,\cdots,N\}$，给定任意一个输入向量 $\boldsymbol{X}^*$，则通过非线性映射到高维特征空间，然后在高维特征空间中进行线性回归得到预报输出值$\hat{y}$，即 RVM 的模型输出表示为

$$\hat{y}=\sum_{i=1}^{N}w_iK(\boldsymbol{X}^*,\boldsymbol{X}_i)+w_0 \tag{3}$$

式中：$K(\cdot,\cdot)$ 为核函数，$\boldsymbol{w}$ 为模型的权值.

令训练样本集为$\{\boldsymbol{X}_i,y_i\}_{i=1}^{N}$，且 $p(\boldsymbol{y}\mid\boldsymbol{X})$ 服从 $N(\boldsymbol{y}\mid\hat{y},\sigma^2)$ 的高斯分布，则相应的训练样本集的高斯似然函数为

$$p(\boldsymbol{y}\mid\boldsymbol{w},\sigma^2)=(2\pi\sigma^2)^{-N/2}\exp\left\{\frac{-\|\boldsymbol{y}-\boldsymbol{\Phi w}\|^2}{2\sigma^2}\right\} \tag{4}$$

式中：$\boldsymbol{y}=(y_1 \quad y_2 \quad \cdots \quad y_N)$；$\boldsymbol{w}=(w_0 \quad w_1 \quad \cdots \quad w_N)$；$\boldsymbol{\Phi}$ 为 $N\times(N+1)$ 核函数矩阵，$\Phi_{mi}=K(\boldsymbol{X}_n,\boldsymbol{X}_n)$，$\Phi_{n1}=1$. 利用极大似然函数法估计 $\boldsymbol{w}$、σ 时会导致严重的过拟合现象[8]，因此，为每个 $\boldsymbol{w}$ 定义高斯先验概率分布函数

$$p(\boldsymbol{w}\mid\boldsymbol{\alpha})=\prod_{i=0}^{N}N(w_i\mid 0,\alpha_i^{-1}) \tag{5}$$

然后基于贝叶斯准则计算权值的后验概率分布，即

$$p(\boldsymbol{w}\mid\boldsymbol{y},\boldsymbol{\alpha},\sigma^2)=(2\pi)^{-(N+1)/2}\mid\boldsymbol{\Sigma}\mid^{1/2}\times \exp\{-(\boldsymbol{w}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})/2\} \tag{6}$$

式中：后验分布的协方差和均值分别为 $\boldsymbol{\Sigma}=(\sigma^2\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\boldsymbol{A})^{-1}$，$\boldsymbol{\mu}=\sigma^2\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y}$，其中 $\boldsymbol{A}=\mathbf{diag}\{\alpha_0,\alpha_1,\cdots,\alpha_N\}$.

通过最大化边缘似然分布函数

$$p(\boldsymbol{y}\mid\boldsymbol{\alpha},\sigma^2)=N(0,\sigma^2\boldsymbol{I}+\boldsymbol{\Phi A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}) \tag{7}$$

即可得超参数估计值，本文采用 EM 算法[8] 内循环迭代估计超参数$\boldsymbol{\alpha}$、σ，其α_i 和σ 的迭代方程分别如下：

$$\alpha_i^{\mathrm{new}}=\frac{1}{\Sigma_{ii}+\mu_i^2} \tag{8}$$

$$(\sigma^2)^{\mathrm{new}}=\|\boldsymbol{y}-\boldsymbol{\Phi\mu}\|^2/(N-\sum_i\gamma_i) \tag{9}$$

式中：$\gamma_i=1-\alpha_i\Sigma_{ii}$.

获得超参数最优值 α_{mp}、σ_{mp}^2，则 $\boldsymbol{w}=\boldsymbol{\mu}_{\mathrm{mp}}$，由于权值最优估计的不确定性，给定任意一个输入值 $\boldsymbol{X}^*$，则描述预测不确定度的均值和方差如下：

$$\boldsymbol{\mu}^*=\boldsymbol{\Phi}(\boldsymbol{X}^*)\boldsymbol{\mu}_{\mathrm{mp}},\sigma_*^2=\sigma_{\mathrm{mp}}^2+\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{X}^*)\boldsymbol{\Sigma}\boldsymbol{\Phi}(\boldsymbol{X}^*) \tag{10}$$

3 模型参数辨识及计算流程

RVM 径流预报模型需要解决两个问题：(1)模型核函数选择；(2)模型参数辨识过程中的目标函数确立. 对于核函数的选择，线性核函数是径向基核函数的特例，特定的 sigmoid 核函数功能上与径向基核函数相同，核函数自身参数的个数太多不利于参数的选择[7]，因此，本文选取径向基核函数为核函数；其次，在模型参数识别过程中，通常选取训练样本的拟合误差最小为模型目标函数，但是该方式下训练误差收敛过程中会出现严重的过拟合现象，如图 1 中拟合曲线 1 所示，训练阶段拟合误差非常小，几乎接近于零，导致优化参数不合理，外推期预报精度非常低. 因此本文在训练过程中考虑具有丰、平、枯年份的检验样本误差来抑制过拟合，即综合考虑训练样本和检验样本的误差建立目标函数，如下式：

$$\min R=\left(\frac{N_2}{N_1+N_2}\right)R_1+\left(\frac{N_1}{N_1+N_2}\right)R_2 \tag{11}$$

式中：R_1、R_2 分别为训练期、检验期的相对误差绝对值的平均值，N_1、N_2 分别为训练期、检验期的样本个数. 其收敛过程如图 1 中拟合曲线 2 所示，曲线 2 的收敛值要比曲线 1 的收敛值大，说明本文目标函数具有抑制过拟合的能力.

最后，利用 RVM 模型进行径流预报的主要步骤如下：

(1)给定参数 m、τ、ε 的合理取值区间，对时间序列进行相空间重构；

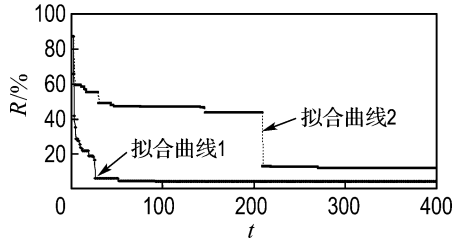


图 1　不同目标函数的训练误差收敛过程

Fig. 1　Training error convergence process of different objective functions

(2)利用重构后的训练样本作为 RVM 的输入进行训练，采用 PSO 算法辨识方法参数，并验证数据序列的混沌特性，在 RVM 内循环中，利用式(8)、(9)迭代估计超参数 α_i、σ，给定一个 $\alpha_{\max}$ 的值，将 $\alpha_i^{\text{new}}>\alpha_{\max}$ 对应的权重值设定为 0，并剔除对应的 $\boldsymbol{X}_i$，所剩余的 $\boldsymbol{X}$ 即为相关向量，对应的权重向量 $\boldsymbol{w}=\boldsymbol{\mu}_{\text{mp}}^{\text{T}}$；

(3)依据优化所得参数为模型的参数取值，给定任意一个输入向量 $\boldsymbol{X}^*$，采用训练好的 RVM 进行计算便可得到预报值的均值 $\boldsymbol{\mu}^*$ 和方差 σ_*^2，预报值服从均值 $\boldsymbol{\mu}^*$ 和方差 σ_*^2 的后验正态概率分布.

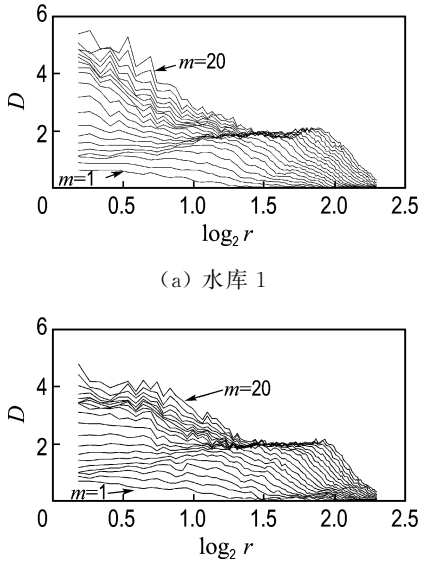
4　应用实例分析

选取南方两水库入库月径流时间序列作为研究实例，分别为水库 1 的 51 a(1953-01～2003-12)和水库 2 的 48 a(1958-01～2005-12)入库月径流时间序列. 水库 1 的控制流域面积为 11.45×10^4 km^2，多年平均月径流量为 1 251 m^3/s，变差系数为 0.717，最大、最小月径流量分别为 5 000、248 m^3/s；水库 2 的控制流域面积为 10.26×10^4 km^2，多年平均月径流量为 1 215 m^3/s，变差系数为 0.856，最大、最小月径流量分别为 5 480、236 m^3/s. 每一个样本序列分成 3 个子样本，对于水库 1，将前 41 a 资料作为模型的训练样本，中间包含丰、平、枯年份的 5 a 资料作为检验样本，与前 41 a 配合确定合理的模型参数，剩余 5 a 不参加确定模型参数，纯粹用于检验确定模型的外推预报能力. 同样，对于水库 2，用前 38 a 的序列点作为训练样本，中间 5 a 作为检验样本，剩余 5 a 序列点作为外推预测样本. 对数据进行规格化处理，采用 PSO 算法优化各方法中的参数，取相对误差的绝对值(E_{mar})、相关系数(R)、确定性系数 R^2 和合格率(定量)作为预报结果的评价指标. 先以水库 1 为例进行计算分析，以水库 2 作进一步的验证.

4.1　实例分析

RVM 模型参数主要有相空间重构参数(m、τ)、所选取的核函数自身参数及模型自动确定的超参数($\boldsymbol{\alpha}$，σ). 给定一个较小的参数区间，进行优化计算，若优化所得参数值为区间端点值，则进一步扩大区间重新计算，直至参数取值在区间范围内为止，即该区间为参数区间，由此确定径向基核函数带宽 ε，混沌时间序列嵌入维数 m、时间滞时 τ 的取值区间分别为[0.1,100]、[1,20]、[1,10]，另外对超参数初始化，取 $\alpha(0)=(0.25,0.25,\cdots,0.25)$，$\sigma_0^2=\text{var}(y)\times0.01$，$\alpha_{\max}=1\times10^5$.

其次，对时间序列进行相空间重构，以重构后的训练样本作为径流模型的输入条件，采用 PSO 优化模型参数(ε、m、τ)，水库 1 结果为(2.172 3，14，4)，水库 2 结果为(4.554，7，6). 根据优化所得时间滞时，分别以嵌入维数 $m=1\sim20$ 绘制 D-$\log_2 r$ 折线斜率图，如图 2 所示，$\log_2 r$ 在 1～2，且 $m>8$ 时饱和关联维数趋于稳定值，即存在显著标度区，从而可以定性判断两水库月径流序列存在混沌特性，并由图 2(a)、(b)可以估计出 $1<D<3$，重构参数 m 满足 $m\geqslant2D+1$ 的条件，说明重构参数是合理的. 此外，在参数内循环过程中，随着超参数的迭代估计，边缘似然分布函数值逐步趋于稳定，如图 3 所示；由图 4 知超参数 σ^2 收敛很快，迭代 3 次后基本达到最优值，因此有的文献也将超参数 σ^2 作为一个固定值进行内循环计算.



(a) 水库 1

(b) 水库 2

图 2　水库 1 和 2 月径流关联维数图

Fig. 2　Correlation dimension of monthly runoff flow of Reservoirs One and Two

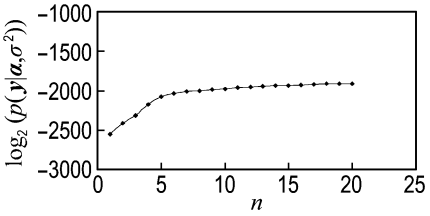


图 3 $\log_2(p(\boldsymbol{y}|\boldsymbol{\alpha},\sigma^2))$的收敛过程

Fig. 3 Convergence process of $\log_2(p(\boldsymbol{y}|\boldsymbol{\alpha},\sigma^2))$

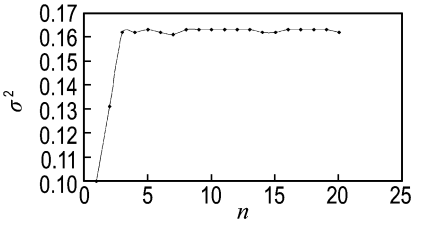


图 4 超参数 σ^2 的收敛过程

Fig. 4 Convergence process of hyper-parameter σ^2

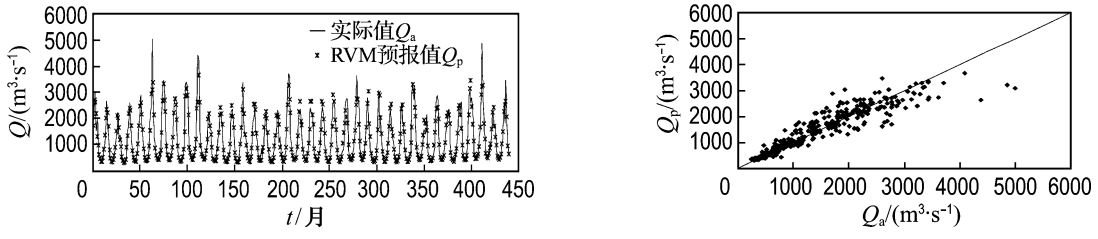
分析模型的模拟、检验、外推预报精度，并将本文模型(RVM)与应用较为广泛的最小二乘支持向量机模型(LSSVM)，以及未考虑相空间重构的($m=12,\tau=1$)相关向量机模型(RVM*)和自动回归滑动平均模型(ARMA(5,6))进行对比分析. 计算结果列于表 1 中，RVM 方法对于训练期、检验期和外推预测期的预报结果见图 5～8.

总体而言，由表 1 知，考虑相空间重构进行预报时比未考虑相空间重构时，RVM 获得比单一方法更优的预报精度，与 LSSVM 和 ARMA(5,6)的计算结果相比较，RVM 的评价指标值均优于其相应值，说明 RVM 具有较好的预报性能. 按照《水文情报预报规范》(SL 250—2000)标准将径流量划分为枯、偏枯、平、偏丰、丰 5 个级别，对高流量(包括偏丰和丰流量)精度进行了定量、定性比较分析，结果如表 2 所示. RVM 在训练期、检验期及外推预测期的平均绝对相对误差分别比 LSSVM 和 ARMA 的相应值要小，但同时其预报精度均比预报总体时相应值低，以多年变幅的 20%为许可误差，比较分析知，其定量合格率较本

表 1 水库 1 月流量不同方法预测精度

Tab. 1 Prediction accuracy of monthly flow of Reservoir One resulting from various methods

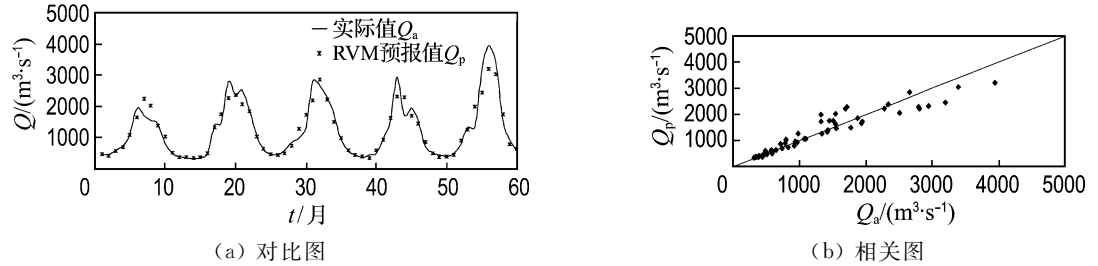
方法	训练期				检验期				外推预测期			
	E_{mar}/%	R	合格率/%	R^2	E_{mar}/%	R	合格率/%	R^2	E_{mar}/%	R	合格率/%	R^2
RVM	12.34	0.94	82.2	0.88	11.92	0.95	83.3	0.91	13.59	0.93	76.7	0.81
RVM*	15.62	0.92	75.4	0.83	17.58	0.89	68.3	0.79	14.93	0.92	66.7	0.75
LSSVM	13.76	0.93	76.1	0.86	13.37	0.93	76.7	0.85	14.31	0.92	73.3	0.77
ARMA(5,6)	18.27	0.91	62.2	0.72	17.75	0.89	61.7	0.72	17.07	0.93	65.0	0.70



(a) 对比图 (b) 相关图

图 5 训练期流量实测值与 RVM 预报值对比图及相关图

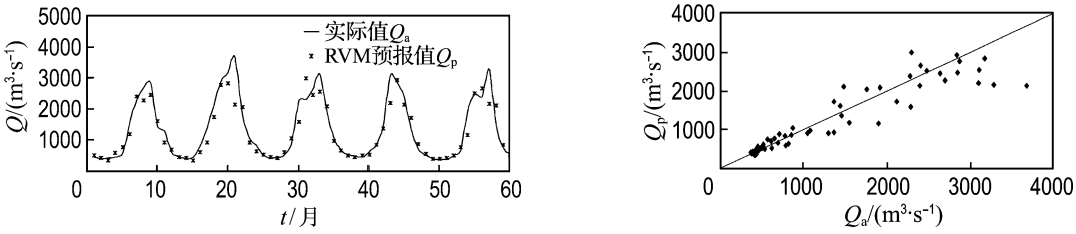
Fig. 5 Comparison and scatter plot between observed flow and predicted flow by RVM during training period



(a) 对比图 (b) 相关图

图 6 检验期流量实测值与 RVM 预报值对比图及相关图

Fig. 6 Comparison and scatter plot between observed flow and predicted flow by RVM during test period



(a) 对比图　　(b) 相关图

图 7　外推预测期流量实测值与 RVM 预报值对比图及相关图

Fig. 7　Comparison and scatter plot between observed flow and predicted flow by RVM during validated period

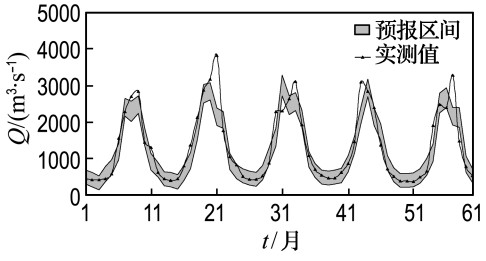


图 8　外推预测期实测流量与 RVM 预报区间对比图

Fig. 8　Comparison between observed and predicted interval hydrograph during validated period

文所列其他方法有所提高. 为提供更为充分的预报信息，本文对比分析了三阶段高流量的定性预报合格率，除了在检验期 RVM 和 LSSVM 的合格率相同以外，其余两阶段 RVM 均获得比其他方法更高的定性预报合格率，同样说明 RVM 具有较强的高流量预报能力.

进一步考虑径流预报的不确定性，以预报值的均值和方差为预报的后验概率分布函数来描述预报值的不确定性，并讨论了发生概率为 80% 的区间预报，其区间预报结果及实测流量过程如图 8 所示. 由图 8 知，中低流量预报区间基本上可以包住实测流量，高流量区间上下限值对应的级别能够预报出实测值对应的级别，概率区间预报是可靠的.

4.2　实例验证分析

水库 2 的统计参数与水库 1 基本相同，但是变差系数较大，数据序列平稳性相对较差，在同样可行条件下，对水库 2 进行了计算，其预报结果的评价指标列于表 3. 由表 3 知，RVM 较 LSSVM 和 ARMA(6,6)模型具有较高的预报精度，验证说明了本文模型的有效性.

表 2　水库 1 高月流量不同方法预测精度

Tab. 2　Prediction accuracy of high monthly flow of Reservoir One resulting from various methods

方法	训练期			检验期			外推预测期		
	E_{mar}/%	定量合格率/%	定性合格率/%	E_{mar}/%	定量合格率/%	定性合格率/%	E_{mar}/%	定量合格率/%	定性合格率/%
RVM	14.27	75.4	92.9	14.92	73.9	82.6	16.53	70.8	83.3
LSSVM	18.24	71.6	87.8	18.94	67.8	82.6	18.13	66.6	79.1
ARMA(5,6)	19.03	62.1	74.3	22.41	56.5	70.0	19.32	62.5	76.6

表 3　水库 2 月流量不同方法预测精度

Tab. 3　Prediction accuracy of monthly flow of Reservoir Two resulting from various methods

方法	训练期				检验期				外推预测期			
	E_{mar}/%	R	合格率/%	R^2	E_{mar}/%	R	合格率/%	R^2	E_{mar}/%	R	合格率/%	R^2
RVM	13.00	0.938	81.7	0.86	11.17	0.959	80.0	0.89	14.62	0.907	75.0	0.84
LSSVM	15.68	0.921	76.6	0.85	12.64	0.943	73.3	0.79	15.65	0.901	71.7	0.76
ARMA(6,6)	18.74	0.903	60.7	0.71	18.98	0.900	56.7	0.71	19.15	0.885	51.7	0.67

5　结　论

(1)将混沌技术与相关向量机结合建立径流预报模型，采用 PSO 算法辨识模型参数，优化所得重构参数满足混沌理论条件，耦合方法比单一方法的预报精度有所提高，并对总体和高流量值进行分析，取得比 LSSVM 和 ARMA 模型更优的预报精度，说明本文模型的有效性.

(2)相关向量机为概率模型，能够定量地、以概率分布的形式描述径流预报不确定性，并给出

指定发生概率下的区间预报.

(3)在进行中长期径流预报应用时,相关向量机模型的不足之处是模型参数和样本序列均以正态概率分布函数进行推理,但从模型计算的结果来看可用于中长期径流预报,下一步将以P-Ⅲ型概率分布函数进行模型研究.

参考文献:

[1] SIVAKUMAR B. Chaos theory in hydrology important issues and interpretations [J]. **Journal of Hydrology**, 2000, **227**:1-20

[2] SIVAKUMAR B, JAYAWARDENA A W, FERNANDO T M K G. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches [J]. **Journal of Hydrology**, 2002, **265**:225-245

[3] 王 文,许武成. 对水文时间序列混沌特征参数估计问题的讨论[J]. 水科学进展, 2005, **16**(4):606-610

[4] YU X Y, LIONG S Y, BABOVIC V. EC-SVM approach for real time hydrologic forecasting [J]. **Journal of Hydroinformatics**, 2004, **6**(3):209-223

[5] SIVAKUMAR B. Nonlinear determinism in river flow prediction as a possible indicator [J]. **Earth Surface Processes and Landforms**, 2007, **32**:969-979

[6] LIONG S Y, SIVAPRAGASAM C. Flood stage forecasting with SVM [J]. **Journal of the American Water Resources Association**, 2002, **38**(1):173-186

[7] 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

[8] TIPPING M E. The relevance vector machine [J]. **Advances in Neural Information Processing System**, 2000, **12**:652-658

[9] AGARWAL A, TRIGGS B. 3D human pose from Silhouettes by relevance vector regression [J]. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 2004, **2**:882-888

[10] BOWD C, MEDEIROS F A, ZHANG Zuo-hua, *et al*. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements [J]. **Investigative Ophthalmology & Visual Science**, 2005, **46**:1322-1329

[11] CHEN S, GUNN S R, HARRIS C J. The relevance vector machine technique for channel equalization application [J]. **IEEE Transactions on Neural Networks**, 2002, **12**(6):1529-1532

[12] KENNEDY J, EBERHART R C. Particle swarm optimization [C] // **Proceedings of IEEE Conference on Neural Networks**. Piscataway: IEEE Press, 1995: 1942-1948

[13] KANTZ H, SCHREIBER T. **Nonlinear Time Series Analysis** [M]. Cambridge: Cambridge University Press, 1997

Research on mid- and long-term runoff forecast model with relevance vector machine

SHI Yu-zhi[1,2], PENG Yong*[1], ZHOU Hui-cheng[1]

(1. School of Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China;
2. Water Research Institute of Shandong Province, Jinan 250013, China)

Abstract: Due to the superior forecasting performance, relevance vector machine (RVM) was applied to mid- and long-term runoff forecasting, and based on the phase space reconstruction, the runoff relevance vector machine forecasting model was established. Firstly, the runoff time series was reconstructed in the phase space, and the reconstructed series was as the proposed model input; Secondly, the particles swarm optimization (PSO) algorithm was applied to identifying the model parameters and chaotic properties of time series. The EM algorithm was used to estimate hyper-parameters in the inherent cycle, RVM was compared with widely used least squares support vector machine (LSSVM) and auto-regressive moving average model (ARMA). The test results show that the model has good generalization ability; Finally, in terms of the uncertainty of hydrological processes, the scheduling staffs consider the uncertainties in forecasting, and quantitatively estimate the risks and benefits in decision-making based on the uncertainty of RVM output values and the probability forecast interval.

Key words: phase-space reconstruction; relevance vector machine; long-term runoff forecast; PSO algorithm