

DOI :10.3876/j.issn.1000-1980.2009.03.005

滑动窗口二次自回归模型在径流预报中的应用

任 政 郝振纯

(河海大学水文水资源与水利工程科学国家重点实验室,江苏 南京 210098)

摘要:为提高径流预报精度,采用滑动窗口二次自回归模型进行径流中长期预报,提出自适应优化、平稳滑动窗口和均方根误差最小的综合选取模型参数的方法.实例应用结果表明:该方法不仅提高了模型的预报精度,而且保证了模型的稳定性;与人工神经网络模型相比,滑动窗口二次自回归模型的 1 步预报具有更高的精度,可用于中长期径流预报.

关键词:中长期径流预报;二次自回归;滑动窗口;自适应

中图分类号:TV121⁺.4;P338⁺.2 文献标识码:A 文章编号:1000-1980(2009)03-0267-04

中长期径流预报由于具有较长的预见期,使人们在解决防洪与抗旱、蓄水与弃水以及各部门用水之间的矛盾时,能够及早采取措施进行统筹安排,以便获取最大效益.中长期径流预报主要采用成因分析方法和数据挖掘技术^[1-2],由于径流过程受太阳辐射、地心引力、下垫面和人类活动等因素影响,基于成因分析的中长期预报往往比较困难,而径流的历史数据通常隐藏未来形成径流的规律,因此,数据挖掘技术成为研究中长期径流预报的有效途径.目前应用较广的模型有自回归模型(auto-regressive model,ARM)^[3]、灰色模型(grey model,GM)^[4-5]、人工神经网络(artificial neural network,ANN)模型^[6]和支持向量机(support vector machine,SVM)模型^[7-8].滑动窗口二次自回归(moving windows quadratic auto-regressive,MWQAR)模型在传统的自回归模型基础上构造二次项,假设在滑动的窗口内非线性系统用线性结构近似,根据输入向量及模型参数做出预测.该模型在径流预报应用中的研究很少,李爱国等^[9-10]已将该模型成功应用于 Hénon 混沌时间序列和股票交易价格预测中.

MWQAR 模型在径流的预报中需要解决 3 方面的问题:一是数据的预处理问题(文献[9-10]应用的混沌时间序列中,中长期的月径流往往存在一定的周期性规律);二是相关参数的辨识(文献[9-10]通过预先选定窗口大小以及模型的一次项和二次项的阶数,带有随机性和任意性);三是模型的稳定性.针对上述问题,本文采用半正弦函数方法过滤中长期月径流数据,通过自适应优化选择参数^[11];采用普适似然不确定性估计(generalized likelihood uncertainty estimation,GLUE)理论^[12-13],设定似然函数的阈值选择参数集,分析窗口不同滑动步伐的误差隶属度与参数集中顺序元素的一致性,选择最优稳定的预报模型.为验证模型的可行性,以新疆玛纳斯河肯斯瓦特站月径流实测资料作为实例应用,并与 ANN 模型结果进行对比.

1 MWQAR 模型

设径流时间序列训练样本 $x_t(t=1,2,\dots,n)$ 二次型自回归模型描述为^[9]

$$\hat{x}_{t+T}=x_t w_1+x_{t-1} w_2+\dots+x_{t-(p-1)} w_p+x_t^2 v_1+x_{t-1}^2 v_2+\dots+x_{t-(q-1)}^2 v_q+e \quad (p>0,q>0) \quad (1)$$

式中: p,q ——一次项和二次项的阶数; e ——零均值白噪声; T ——预报期.选择一个合适的窗宽 m ,假设二次型自回归模型在窗口内能够反应一个非线性动力系统,且随着时间的推移,窗口也向前滑动,则式(1)演化为以下模型:

$$X=\Phi(Z,\theta)+\Theta \quad (2)$$

其中

$$Z=(Z_{t-T},Z_{t-T-1},\dots,Z_{t-T-(m-1)})^T \quad X=(x_t,x_{t-1},\dots,x_{t-(m-1)})^T$$

式中: Φ ——式(1)中的非线性算子; θ —— $p+q$ 的系数列向量; Θ ——零均值白噪声列向量; Z ——窗口内的输入向量; X ——窗口内的输出向量.式(2)称为 MWQAR 模型.

在式(2)中已知参数 p, q 和 m 的条件下,系数 θ 的最小二乘估计公式表示为

$$\theta = (Z^T Z)^{-1} Z^T X \tag{3}$$

通过构造向量

$$Z_t = (x_t, x_{t-1}, \dots, x_{t-(p-1)}, x_t^2, x_{t-1}^2, \dots, x_{t-(q-1)}^2)^T$$

则第 t 时刻 T 步预报为

$$\hat{x}_{t+T} = Z_t^T \cdot \theta \tag{4}$$

构建一个优良的 MWQAR 模型作预报,关键是确定参数 p, q 和 m ,以及在窗口滑动过程中模型的稳定性分析,记参数空间为 Ω .

2 模型参数识别和优选

2.1 参数选择

合理选择参数是提高模型预报精度的重要途径,本文采用自适应参数优化的方法选择 MWQAR 模型参数.给定参数 p, q 和 m 的取值范围,先验分布采用均匀分布.由于不同的参数组合依赖的训练样本数是不相同的,相同的参数组合窗口的滑动也会导致训练样本不相同,因此,根据选定的参数组合,从训练样本起点开始逐步滑动窗口,建立如式(2)所示的预报模型并对窗口外的第一个值进行预报,直到训练样本的最后一个值.采用 GLUE 理论,取效率系数 $N_S^{[12]}$ 作为似然目标函数,设定临界似然函数值 α ,如果 $N_{S_i} < \alpha$,取 $N_{S_i} = 0$,表示这些参数组不能表达模型的功能特征.将高于临界值的参数组的似然函数值按照大小重新排序,计算相应的累计概率 P 并得到参数选择空间 Ω .

2.2 参数优选

参数的优选通过率定样本进行选择,将 Ω 中的元素通过率定样本,逐步滑动窗口,计算各个窗口下 Ω 中元素对应的均方根误差,将最小的均方根误差赋予隶属度 1,得到 Ω 中元素对应的隶属度,按照隶属度的大小将 Ω 中的参数进行重新排序,从而得到各个窗口下的 Ω 中元素的顺序表.以给定某一概率 P 为界,在多个顺序表中寻找公共元素,如果存在元素个数超过 1 个,则缩小 P ;如果不存在,则增大 P ,直到找到唯一的公共元素为止,则以此元素作为最优选择的参数组合.简单的数学描述如下:

设 Ω 中的元素为 a_i ,对应的序号 h_i 的累计概率为 $P_i (i = 1, 2, \dots, l)$,率定样本为 Y ,将 Y 划分为不同的滑动窗口,分别记为 $Y_k (k = 1, 2, \dots, K)$;均方根误差记为 R ;隶属度记为 μ .那么,存在 $R_{i,k} = f(a_i, Y_k)$,取 $\mu_{i,k} = R_{i,k} / R_{\min,k}$,按照 $\mu_{i,k}$ 的大小将 h_i 重新排序为 $h_{i,k}$,并且记 $H_k = \{h_{i,k} | i = 1, 2, \dots, l\}$, h_i 的序号不变,仅是位置发生改变, $\forall P$ 满足 $h_i \in \bigcap H_k$,且 h_i 是唯一的, P 从小到大选取.

根据所求的 h_i ,选择对应的 a_i 作为参数的选择.通过以上步骤选择参数,模型达到了 2 个要求:一是模型对训练数据具有较高的精度;二是在不同的滑动窗口下模型具有相对平稳的滑行速度.

3 MWQAR 模型的应用

以新疆玛纳斯河肯斯瓦特站月径流实测资料(1954 年 1 月~1995 年 12 月)为研究对象,说明 MWQAR 模型在中长期径流预报中的应用.玛纳斯河流域的径流以降雨补给为主,兼有融雪补给,观测站年平均流量 $39.86 \text{ m}^3/\text{s}$,年来水量 12.57 亿 m^3 ,属干旱地区. MWQAR 模型选取前 420 个月径流数据训练模型,其后 36 个月径流数据率定模型,最后 48 个月径流数据测试模型;ANN 模型选取前 456 个月径流数据训练模型,其后 48 个月径流数据测试模型.

3.1 输入数据预处理

数据预处理是 MWQAR 模型预报的前期工作,其目的是过滤掉数据中隐含的一些周期性规律.将径流数据进行快速傅立叶变换(fast fourier transform, FFP)^[14],图 1 结果显示数据具有 3 个明显的周期,相应周期分别为 12, 3 和 4,取最大周期为 12.由于径流年内分配主要集中在 6~8 月,因此采用半正弦函数对原始数据进行过滤,相应的过滤函数表达如下:

$$\hat{y}_i = A \sin(0.083\pi r_i) + b \tag{5}$$

用训练数据对式(5)进行训练,计算参数 $A = -118.52$, $b = 114.65$. 图 2 显示出过滤后数据振幅下降,消除了月径流的周期性规律.

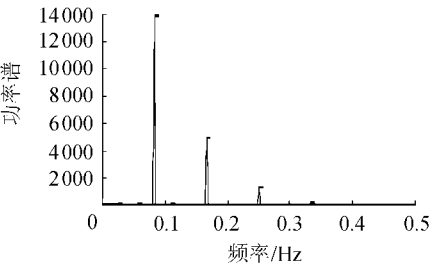


图 1 训练数据的功率谱

Fig. 1 Power spectra of training data

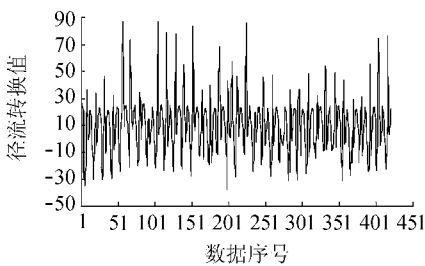


图 2 训练数据预处理

Fig. 2 Pretreatment of training data

3.2 模型参数选择及预报结果

用 MWQAR 模型进行水文中长期径流预报,关键是确定一次项阶数 p 、二次项阶数 q 以及滑动窗宽 m . 结合文献 [9-10] 以及本文的研究对象,选择以上 3 个参数的取值范围分别为 $p \in [5, 15]$, $q \in [5, 15]$, $m \in [100, 200]$,设置参数自适应步长为 1. 结合文献 [12] 选取 $\alpha = 0.47$,计算得到 Ω 的元素为 314 个,取 $P = 0.1$,最优稳定参数为(12, 11, 197). 按照以上参数取值进行 MWQAR 模型的 1 步(MWQAR1)和 2 步(MWQAR2)预报,同时选用 3 层前馈反向传播 ANN 模型进行比较,其反向传播学习算法采用的是 SCG 算法,并通过试验确定最优的 ANN 隐层数为 3. 为比较 MWQAR 模型与 ANN 模型的预报效果,2 种模型均采用预处理后的数据. 图 3 为测试段的实测值和 2 种模型的预报值,表 1 为 2 种模型的预报结果的相关系数和均方根差,结果显示 1 步 MWQAR 模型比 ANN 模型预报精度有一定的提高.

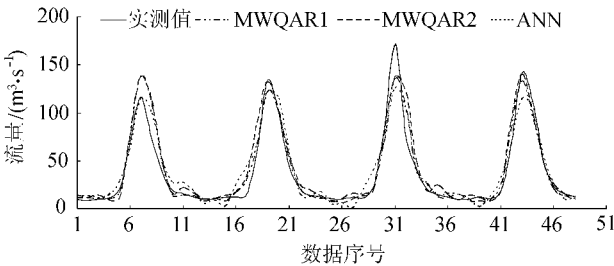


图 3 模型测试段 ANN 模型与 MWQAR 模型预报值

Fig. 3 Predicted results of model test section by MWQARM and ANNM

表 1 MWQAR 模型和 ANN 模型的预报精度		
Table 1 Prediction precisions of MWQARM and ANNM		
模 型	相关系数	均方误差
MWQAR1	0.9525	10.5413
ANN	0.9336	12.5945
MWQAR2	0.9171	12.6462

4 结 论

将 MWQAR 模型应用于中长期径流预报中,在文献 [9] 的基础上,提出自适应、平稳滑动窗口和均方根误差最小的参数选取方法,选取的参数不仅提高了模型的预报精度,而且保证了模型的稳定性.通过对径流的频谱分析,巧妙地利用半正弦函数规避数据的周期性规律,大大降低了序列的波动范围.但是,该方法对于周期的消除并不具有普适性,这与研究区径流的年内分配特点关系密切.同时,参数的选取和临界似然函数目标值的选取具有一定的主观性,如果选取得不合适,可能丢掉更好的参数组合.如何确定它们的选择原则,从理论上做出一个逻辑的证明,将有助于该方法的理论发展和实际应用.

参考文献:

[1] 杨旭,栾继虹,冯国章.中长期水文预报研究评述与展望[J].西北农业大学学报,2000,28(6):203-207.
[2] 王文,马骏.若干水文预报方法综述[J].水利水电科技进展,2005,25(1):56-60.
[3] 吴学生.应用自回归模型进行长期水文预报[J].东北水利水电,2005,23(1):41-45.
[4] 徐凤才,杨杰.灰色系统理论在中长期洪水预报中的应用[J].东北水利水电,2004,22(7):10-11.
[5] 路振广.改进的灰 GM(1,1)模型在灌区水文年型中长期预报中的应用[J].灌溉排水学报,2005,24(5):39-41.

- [6] 鞠琴,郝振纯,余钟波,等.基于人工神经网络的降雨径流模拟研究[J].辽宁工程技术大学学报,2007,26(6):940-943.
- [7] VAPNIK V N.统计学习理论的本质[M].张学工,译.北京:清华大学出版社,2000.
- [8] 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.
- [9] 李爱国.滑动窗口二次自回归模型预测混沌时间序列[J].系统工程理论与实践,2004,24(10):104-109.
- [10] 李爱国,覃征.滑动窗口二次自回归模型预测非线性时间序列[J].计算机学报,2004,27(7):1004-1008.
- [11] 刘涵,刘丁,李琦.基于支持向量机的混沌时间序列非线性预测[J].系统工程理论与实践,2005,25(9):94-99.
- [12] McMICHAEL C E, HOPE A S. Distributed hydrological modeling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation[J]. Journal of Hydrology, 2006, 317: 307-324.
- [13] 黄国如,谢河海.基于GLUE方法的流域水文模型的不确定分析[J].华南理工大学学报,2007,35(3):137-142.
- [14] 代涛,邵东国,黄显峰.基于混沌理论的水资源时空变异性[J].武汉大学学报:工学版,2007,40(5):15-19.

Application of moving window quadratic auto-regressive model in runoff prediction

REN Zheng, HAO Zhen-chun

(State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
Hohai University, Nanjing 210098, China)

Abstract: In order to improve the precision of prediction of runoff, the moving window quadratic auto-regressive model was adopted for the prediction of medium and long-term runoff. A comprehensive method for choosing the model parameters is put forward based on self-adapting optimization, smoothly moving windows and the minimum root mean square error. The results show that the present method not only improves the prediction precision, but also ensures the model stability. Compared with the artificial neural network model, the present model has higher prediction precision for its first step and can be applied in the prediction of medium and long-term runoff.

Key words: medium and long-term runoff prediction; quadratic auto-regression; moving window; self-adapting

· 简讯 ·

第2届全国土工合成材料防渗排水学术研讨会将在南京召开

土工合成材料防渗排水技术广泛应用于水利水电、道路交通、港口航道、建筑市政、环保生态等工程领域,经过多年的工程实践和科学研究,取得了长足发展.为了总结成功经验,传播推广科研成果,促进土工合成材料防渗排水技术在工程及科研领域的全面提升,中国土工合成材料工程协会、国际土工合成材料学会中国委员会定于2009年10月24~26日在江苏省南京市(河海大学校内)举办“第2届全国土工合成材料防渗排水学术研讨会”.会议将由相关专家作专题报告,开展学术交流与研讨,参观相关工程,正式出版学术会议论文集,成立中国土工合成材料工程协会防渗排水专业委员会.

会议交流议题与征文内容如下:(a)土工合成材料防渗排水理论;(b)防渗排水结构设计原理;(c)防渗排水新材料产品及其物理力学性能;(d)防渗材料拼接工艺与铺设工艺及装备;(e)排水反滤材料施工工艺及装备;(f)抗老化防渗材料的工程应用性态;(g)软土中排水带的加筋效应;(h)防水力侵蚀工程中(河海岸防护、水土保持等)的防护技术;(i)围海造陆(造库)工程中的防渗排水防护技术;(j)防渗排水防护工程实例.论文提交时间:2009年5月31日前提交论文摘要(200~400字);2009年7月31日前提交论文全文电子稿(论文格式另行通知).

会议秘书处联系方式:

电话:025-83786883,15950576015

电子邮箱:haimin-wu@hotmail.com

联系人:吴海民

(本刊编辑部供稿)