

基于灰色遗传 BP 神经网络的校园区间需水预测研究

杨利纳, 李文竹, 刘心

(河北工程大学, 河北 邯郸 056038)

摘要: 水资源预测是城市安全用水的基础保障,而校园用水预测是城市用水规划和管理的组成部分。针对校园用水受很多因素影响产生的不确定性,提出了基于灰色遗传 BP 神经的校园用水预测模型。模型对校园用水的数据进行灰色关联分析,并加入遗传算法去优化 BP 神经网络,经过残差计算,输出区间的预测值。运用该模型可以充分提取小样本信息,解决神经网络无法自动寻优的问题。通过 Matlab 对校园的用水区间数据进行仿真,得出的结果显示,预测的数据和实际数据基本吻合,其仿真精度可以达到 90.32%,验证了该方法的可行性,此预测方法有一定的借鉴意义。

关键词: 灰色关联分析; 遗传算法; BP 神经网络; 区间需水预测; 残差

中图分类号:TV213.4

文献标识码:A

文章编号:1672-643X(2019)03-0133-06

On campus interval water demand prediction based on grey genetic BP neural network

YANG Lina, LI Wenzhu, LIU Xin

(Hebei University of Engineering, Handan 056038, China)

Abstract: Water resource prediction is the basic guarantee of urban safe water use, while campus water use prediction is the basis of urban water use planning and management. Aiming at solving the uncertainty of campus water consumption caused by several factors, a prediction model of campus water consumption based on grey genetic BP neural network is proposed. The model carries out grey relational analysis on the data of campus water consumption, and adds genetic algorithm to optimize BP neural network. After residual calculation, the predicted value of the output interval is obtained. This model can fully extract information from small samples and add genetic optimization neural network to solve the problem that neural network cannot automatically optimize. The results of Matlab simulation of campus water use interval data show that the predicted data and actual data are basically consistent with a simulation accuracy 90.32%, and it verified the feasibility of the method, the prediction method has some reference significance.

Key words: grey relational analysis; genetic algorithm; BP neural network; interval water demand prediction; residue

1 研究背景

从宏观而言,准确有效的水资源预测可为社会和经济的发展、粮食和生态环境安全提供重要保障^[1]。具体到校园的水资源管理,由于学校校园人口密集,需水量大,所以做好校园的定量用水预测,可以及时发现校园用水异常,以此来合理地下用水

水指标^[2]。

关于需水量研究的方法,国内外构建了许多与需水预测相关的模型,例如 ARIMA 模型^[3]、灰色建模模型^[4]、神经网络模型^[5]。除了上述的方法,有很多学者也将多种预测方法组合在一起。例如,宋帆等^[6]利用聚类方法和灰色关联分析的方法结合起来,对吉林省的用水量进行了预测;查木哈等^[7]

收稿日期:2018-08-06; 修回日期:2018-12-02

基金项目:国家自然科学基金项目(61440001);教育部新世纪优秀人才支持计划项目(NCET-13-0770);河北省高等学校高层次人才科学研究项目(GCC2014062)

作者简介:杨利纳(1993-),女,河北邯郸人,硕士研究生,主要从事智慧节水、需水预测研究。

通讯作者:刘心(1980-),男,吉林敦化人,博士,教授,博士生导师,主要从事智慧节水、宽带通信网络研究。

利用BP神经网络的双层隐层模型,对老哈河流域的用水量进行了预测。但将灰色关联分析和神经网络结合起来的比较少,因此本文尝试将两种算法结合,另外,在神经网络中加入遗传算法优化,不容易陷入局部极小的问题^[8]。

本文建立了以灰色关联分析方法和遗传算法的遗传神经网络区间预测模型,该模型首先采用灰色关联分析方法,通过设置灰色关联度阈值确定需水量主要影响因子,用于提取特征变量,从而降低冗余输入变量的复杂性;其次在神经网络中加入遗传算法优化,不容易陷入局部极小的问题;最后提出了区间需水预测^[9],根据模型先得出点的预测值,再根据点预测和实际值两者的残差值来估计预测区间的上界和下界,区间预测能准确地预测用水量在未来时段的波动范围。此模型可以克服校园区间水预测比较复杂、信息比较单一的缺点,从而避免了灰色关联分析方法及模型在非线形预测方面的缺陷,具有良好的泛化能力和训练效果。通过与传统神经网络方法相比,该模型应用于校园需水预测是可行的。

2 数据来源和研究方法

2.1 灰色关联分析的方法建模

目前最常用的数学方法是对校园数据用水进行灰色关联分析,对影响校园用水的数据进行处理分析,从而确定各个影响因素的权重^[10]。影响校园用水有很多复杂的因素,例如节假日、当日最高气温、最低气温、天气等,本文以天为单位采集了2018年4月份共30d的校园数据,通过灰色关联分析来对校园用水数据进行处理,求出关联系数和关联度,通过比较大小进行排序,得出影响校园用水的主要因素,然后将其作为输入层,输入到神经网络进行校园用水的预测。

2.1.1 原始数据的处理 用灰色关联方法去分析系统的发展趋势,该方法可以在信息贫乏、样本小的系统中提取出对系统指标影响较大的因素,解决计算量过大、样本不服从一定的概率分布、计算结果会出现量化结果和定向结果不同的缺点,本文用区间值化处理方法处理校园原始用水数据^[11]。

2.1.2 灰色关联系数及关联度求解 关联程度可以反映数列形状,校园用水的灰色关联的系数为:

$$\theta_{mn} = \frac{\rho\Delta_{n\max} + \Delta_{n\min}}{\rho\Delta_{n\max} + \Delta_{mn}(i)} \quad (1)$$

式中: $\Delta_{n\max}$ 和 $\Delta_{n\min}$ 分别为所有校园用水比较数列中的最大值和最小值; ρ 为分辨函数; m 为参考序列,

即校园用水量; n 为比较序列,即当日最高气温、最低气温、节假日和天气4个序列; i 取值为30,表示2018年4月份的30d; $\Delta_{mn}(i)$ 为参考序列 m 与比较序列 n 第 i 个点的绝对差值。

关联度计算公式如下:

$$\lambda(x_m, y_n) = \frac{1}{s} \sum_{i=1}^s \theta_{mn}(i) \quad (2)$$

式中: s 为参考序列长度,本文取值为30; $\theta_{mn}(i)$ 为参考序列 m 与比较序列 n 第 i 个值的关联系数; $\lambda(x_m, y_n)$ 为 x 曲线上参考序列 m 与 y 曲线上比较序列 n 对应的关联度。

2.1.3 关联度排序分析 根据数值代表节假日、1天中最高和最低气温,将它们排顺序。关联度描述了参考序列和比较序列相对变化,通俗来讲就是如果两者关联度大,则校园用水和影响因素变化基本一致,反之,关联度小^[12]。

2.2 遗传算法对神经网络的优化

非线性问题用遗传算法解决,进而得到网络的连接权,步骤如下:

(1)采用BP神经网络3层结构确定其权、阈值的长度。

利用训练样本对输入的神经网络进行训练,权值和阈值长度确定后,组成向量^[13]。

(2)确定遗传算法的编码方式及长度和种群初始化。

因为参数比较多,采用实数进行编码,可以提高准确性,使运算更加方便快捷。其中每个实数串都包含了网络的权值和阈值的维数即为编码的长度。

(3)用BP神经网络中的误差定义适应度函数:

$$F(\Psi, \omega, \theta, r) = \frac{1}{\sqrt{\sum_{t=1}^{N_1} \sum_{s=1}^M [y_t(s) - \hat{y}_t(s)]}} \quad (3)$$

式中: Ψ 为BP神经网络中的误差值; ω 、 θ 、 r 分别为神经网络的输入层、隐含层和输出层; N_1 、 M 分别为权值、阈值节点数; $y_t(s)$ 为神经网络的期望输出; $\hat{y}_t(s)$ 为实际输出; t 为需要优化连接权值的1个节点; s 为需要优化阈值的1个节点。

(4)用适应度来进行评测,得出适应度最高的个体在群体中保留下来。

(5)完成交叉和变异操作,则新一代种群由此生成。

(6)得到适应度最高的个体,从而得出神经网络最优的权值和阈值。

2.3 预测方法神经网络模型

当研究一个系统时,将影响因素作为神经网络

的输入,当影响因素比较复杂时,神经网络也变得复杂。文中加入灰色关联分析方法,故降低了网络的性能,减少了计算难度和时间,可以对大量的数据进行处理,其模型计算流程图如图 1 所示。

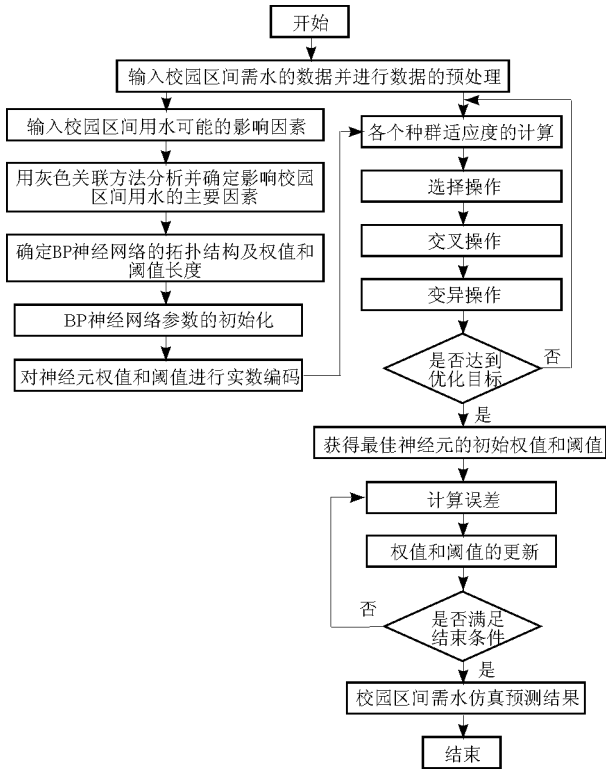


图 1 模型计算流程图

由图 1 可知,通过对数据的分析,将历史数据以及对校园需水影响最大的几个因素作为输入量,输入到神经网络中,各层神经元对影响因素发生作用以后,生成输出量。神经网络的权值和阈值经过遗传算法的优化,得出适应度的数值,以此找出适应度最优的个体。再以输出误差为目标函数,修正误差直到达到要求后,经过训练后的神经网络就可以进行预测。

2.4 预测区间的上下界

由于校园原始数据中没有需水区间的预测上下界,本文通过求出校园用水的点预测值与实际值之间的残差来估计出校园用水预测区间的上界和下界。具体步骤如下:

(1) 根据模型求出的校园用水点预测值和实际的数值进行两者间残差计算。

(2) 用模糊算法^[14]将用水数据样本输入聚类,根据残差大小排序样本。

(3) 计算预测区间的上区间和下区间的方法如公式(4)所示。

$$p = a_m \tag{4}$$

式中: m 的取值若满足 $\sum_{x=1}^m w_{x,y} \geq \frac{\delta}{2} \sum_{x=1}^n w_{x,y}$, 则是预测的上区间,若 m 的取值满足 $\sum_{x=1}^m w_{x,y} < \frac{\delta}{2} \sum_{x=1}^n w_{x,y}$, 预测的则是下区间,其中 a_m 为顺序排好以后样本的残差, $w_{x,y}$ 中 x 为隶属度的值, y 为第 y 个样本; n 为总样本数; p 为预测区间,包括上区间和下区间。

(4) 计算校园用水样本的预测区间。

$$E_j = \sum_{x=1}^f p \tag{5}$$

(5) 计算校园区间上下界,其中 R_y 表示模型输出的点的预测值。

$$T_y = E_y + R_y \tag{6}$$

式(5)和(6)中: E_j 为用水样本的预测区间,包括用水样本预测的上区间和下区间; f 为聚类的总数,其取值为 30; T_y 为校园区间上、下界; R_y 为模型输出点的预测值。

2.5 区间预测优化准则

为了衡量预测区间是否可靠准确,文中引入了预测区间覆盖率 ($PICP$) 和预测区间平均带宽指标 ($PINAW$) 的综合指标 CWC ^[15]。假设规定置信水平为 $100(1 - \alpha)\%$ 。

2.5.1 预测区间覆盖率 ($PICP$)

$$PICP = \frac{1}{N} \sum_{i=1}^N \rho_i \tag{7}$$

式中: N 为样本总数,若预测值在预测区间内,则 $\rho_i = 1$, 否则为 0。若 $PICP$ 很明显小于 $1 - \alpha$, 则说明预测的区间不可靠。

2.5.2 预测区间平均带宽指标 ($PINAW$)

$$PINAW = \frac{1}{NR} \sum_{i=1}^N (m - n) \tag{8}$$

式中: m 和 n 分别为校园用水预测区间的上、下界; N 为样本总数; R 为检验样本目标值的最大值与最小值之间的差值。

2.5.3 CWC 综合评价指标

$$CWC = PINAW + \delta(PICP) e^{\chi(PICP - \eta)} \tag{9}$$

式中: $\eta = 1 - \alpha$, 当 $PICP < \eta$ 时, $\delta(PICP) = 1$, 否则 $\delta(PICP) = 0$; χ 为惩罚参数; CWC 越小区间构建的越好。

3 结果与分析

3.1 影响因子的选择和处理

影响校园区间用水有许多因素,如气象、人文经

济和自然因素^[16]。本文以河北工程大学每日的用水量为研究对象,选取2018年4月份每日的用水量作为输出因子,选取节假日(r_1)、当日最高气温(r_2)、当日最低气温(r_3)和天气(r_4)4个因素作为输入因子构建模型。在2018年4月份的30d中,校园每天的用水量分别与以上4个因素之间的关系如图2所示。

由图2可知,天气因素有晴朗天气、多云天气、下雨天气和阴天。其他的影响校园区间用水的因素还包括当日最高气温和当日最低气温。图中有1个极值点表示用水量最低的时刻,即在4月9日,这天是正常的上课时间,最低气温很低,最高气温也比较低。由此可以看出气温与校园用水量有很大的关系。其影响因素的关联系数经过Matlab仿真的结果如表1所示。

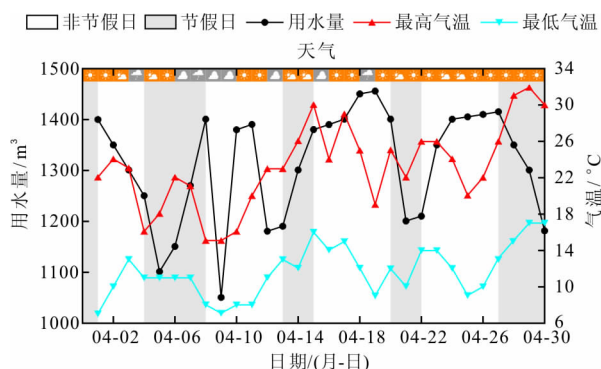


图2 2018年4月份用水量与气温、节假日及天气的关系

表1 影响因素的关联系数

当天最高气温	当天最低气温	节假日	天气
0.8240	0.8906	0.7674	0.5742

由表1可知,影响因素的关联度依次为:当天最低气温 > 当天最高气温 > 节假日 > 天气。当关联度小于0.6,认为两个序列无关,若关联度大于0.8,则两序列相关性很好;在0.6与0.8之间时,两者关联度好^[17]。表1中关联度大于0.6的因素有当日最低气温、当日最高气温和节假日3个因素。

3.2 参数的设置

为了得到更好的预测效果,对BP神经网络和遗传算法进行了最佳的参数的设置,如表2、3所示。

表2 BP神经网络初始参数设置

训练次数	学习率	训练误差
100	0.03	0.05

表3 遗传算法参数设置

种群大小	迭代次数	交叉概率	变异概率
10	50	0.4	0.2

3.3 区间需水量预测结果

本文以河北工程大学校园的用水量为研究对象,所采用的数据来源于河北工程大学节能节水监管平台。首先根据灰色关联分析方法得出校园区间需水量的3个主要影响因素,得出输入节点数为 $m=3$, $n=2m+1$,即隐含层节点数为 $n=7$,可得BP神经网络的结构为3-7-1。此模型用河北工程大学2018年4月份一个月前29d每天24h的数据作为训练值,然后计算出的点预测值见图3。

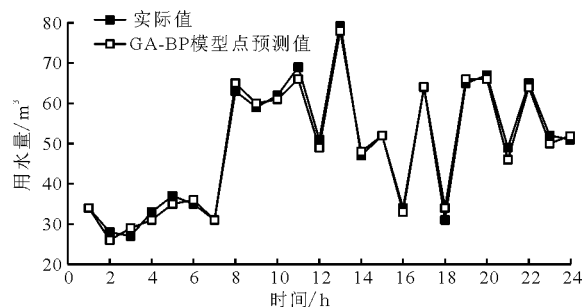


图3 预测模型点预测曲线

从图3可以看出此模型预测值与实测值基本吻合,但是点预测方式一般不能对预测结果做出可靠性描述,因此必须进行模型的区间预测。

3.4 预测模型和BP神经网络的比较

对于校园区间需水预测,本文引入了灰色关联分析方法,找到影响校园区间用水的主要因素,并结合遗传算法优化神经网络解决自动寻优的问题,从而增强了其稳定性和适应能力,通过此方法与BP神经网络预测值进行比较,如图4所示。

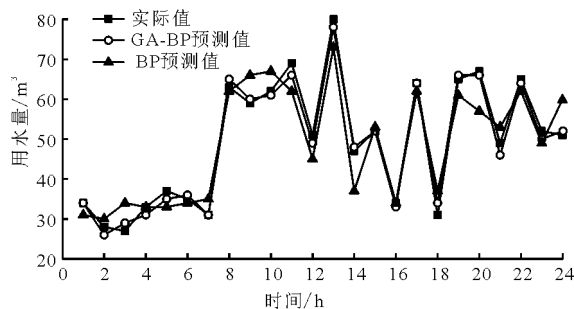


图4 用水量实际值与GA-BP和BP预测值比较图

从图4中可以看出优化后的神经网络不易陷入到极小化的问题,从而使预测结果更精确。

3.5 预测区间上下界结果

本文采用河北工程大学 2018 年 4 月每天的用水量作为样本,假设置信区间为 90%。根据点预测的结果和实际值进行残差计算,可以求出区间的上下界,从而清楚地看到区间预测的结果,见图 5。

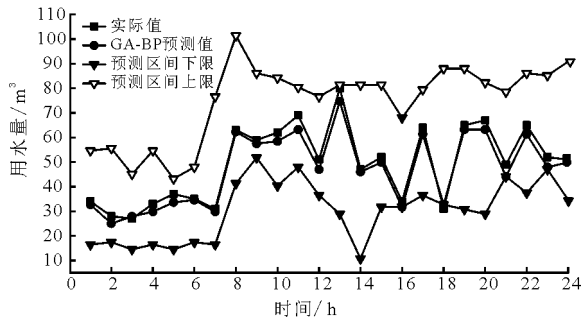


图 5 预测区间上下界及预测结果曲线

从图 5 中可以看出无论预测值还是实际值,基本落入了预测区间上下界区间内,证明该区间预测模型十分有效。

3.6 区间预测的性能指标

区间预测值的可靠性可以用区间覆盖率指标 *PICP*、区间平均带宽指标 (*PINAW*)、综合指标 *CWC* 进行判断。从图 5 中可以看出,实际观测数据中有两个未完全落在预测区间上下限范围内,而训练样本共有 29 个,因而由公式 (7) 可得出 $PICP = 93.10\%$ 。再根据公式(8)、(9)分别得出 *PINAW* 和 *CWC* 的值。3 个指标值的计算结果见表 4。

表 4 区间预测性能指标表

<i>PICP</i>	<i>PINAW</i>	<i>CWC</i>
93.10%	47.84%	0.44

由表 4 可知,预测区间覆盖率 *PICP* 大于 90%,证明此模型建立的预测区间是可靠的。在构建区间模型时,发现 *PICP* 足够大时,*PINAW* 的宽度也越大,而 *PINAW* 越小则 *PICP* 又越小,因而引入 *CWC* 综合指标,*CWC* 的值若足够小,证明本文建立的预测模型是可靠的。

3.7 误差分析

为了验证此算法的精确度,本文引入了误差分析,分析要素包括误差均值、最大误差、最小误差。根据预测值和实际值求出相对误差值^[18]。公式如下:

$$E = \frac{|\gamma_p - \gamma_a|}{\gamma_a} \quad (10)$$

式中: *E* 为相对误差; γ_p 为预测值; γ_a 为真实值。

根据公式 (10) 可以求出误差均值、最大误差、最小误差如表 5 所示。

表 5 误差分析表

误差均值	最大误差	最小误差
3.22	9.68	1.50

由表 5 可知,本文方法所计算的误差均小于 10%,模型的仿真精确度可达到 90.32%,证明此方法预测校园区间需水是十分有效的。

4 结 论

本文结合校园用水数据,用灰色关联法分析出影响校园用水的主要因素,结合遗传算法优化后的神经网络模型可以准确地得到校园用水的点预测值。然后通过输出的点的预测值和实际值的残差来寻找最优的预测区间上下限,可以看出,预测值基本落入了预测区间上下限内。主要结论如下:

(1)通过灰色关联分析方法,找出影响校园用水主要有 3 个因素,即当日最低气温、当日最高气温和节假日,将主要因素作为遗传算法输入层,由计算结果可以看出,预测值曲线和真实值曲线基本一致。引入区间预测,在 90% 的区间置信水平下,预测值基本落入了预测区间上下限内。

(2)为了证明此区间预测的可靠性,引入了预测区间覆盖率、预测区间平均带宽指标和 *CWC* 综合评价指标。通过误差分析,得到相对误差均值为 3.22%、最大误差为 9.68%,证明其仿真精度可以达到 90.32%。

(3)校园需水预测受很多因素影响,为了描述和量化主要因素的影响程度,以河北工程大学为例,引入了上述算法,此模型可以有效地进行校园用水数据的预测,能够及时发现校园用水异常,以此对校园供水进行合理调度,达到节约用水的目的。

参考文献:

[1] 陈建龙,刘永峰,钱 鞠,等. R/S 分析法与 GM(1,1) 灰色模型相结合的鸳鸯池水库入库径流量预测[J]. 水资源与水工程学报,2018,29(5):148-153+158.
 [2] 张子博,刘玉明. 公共建筑节能项目外部性研究——以北京某高校为例[J]. 水资源与水工程学报,2018,29(3):130-137.
 [3] 崔青华,夏井新. 基于 ARIMA-GARCH 模型的城市主干道行程时间时变置信区间预测(英文)[J]. Journal of Southeast University(English Edition),2014,30(3):358-362.

- [4] 杜懿,麻荣永.不同改进灰色模型在广西年用水量预测中的应用研究[J].水资源与水工程学报,2017,28(3):87-90.
- [5] 李晓英,苏志伟,田佳乐,等.基于GRA-MEA-BP耦合模型的城市需水预测研究[J].水资源与水工程学报,2018,29(1):50-54.
- [6] 宋帆,杨晓华,武翡翠,等.灰色关联-集对聚类预测模型在吉林省用水量预测中的应用[J].水资源与水工程学报,2018,29(3):28-33.
- [7] 查木哈,卢志宏,翟继武,等.双隐层BP神经网络模型在老哈河水质预测中的应用[J].水资源与水工程学报,2018,29(2):56-61.
- [8] LI Jichao, ZHAO Danling, GE Bingfeng, et al. A link prediction method for heterogeneous networks based on BP neural network[J]. Physica A: Statistical Mechanics and its Applications, 2018, 495:1-17.
- [9] 王乃珍,王福田.基于灰色区间预测模型的轨道不平顺状态预测[J].铁路计算机应用,2015,24(1):1-3+7.
- [10] 侯珺,周金龙,曾妍妍,等.石河子地区地下水“三氮”空间分布特征及影响因素分析[J].水资源与水工程学报,2018,29(1):1-8.
- [11] SUN Guidong, GUAN Xin, YI Xiao, et al. Grey relational analysis between hesitant fuzzy sets with applications to pattern recognition[J]. Expert Systems With Applications, 2018, 92:521-532.
- [12] 吴梦烟,杨侃,刘朗,等.基于灰色局势决策和组合赋权法的再生水评价[J].水资源与水工程学报,2018,29(3):111-117.
- [13] 高颖,方国华,林泽昕.投影寻踪和BP网络在农村饮用水水质评价中的应用[J].水资源与水工程学报,2017,28(2):115-119.
- [14] 时佳,薛联青,陈新芳,等.基于综合赋权法的叶尔羌河流域水资源承载力可变模糊综合评价[J].水资源与水工程学报,2017,28(5):32-36.
- [15] 杨红霞,蔡昕.基于WDO-PP模型的区域水资源系统与经济社会生态系统协调度评价[J].水资源与水工程学报,2017,28(2):68-75.
- [16] 于增知,于宏兵,郑力燕.松花江流域冷水鱼潜在分布预测模型研究[J].水资源与水工程学报,2017,28(3):48-54.
- [17] 刘佩瑶,郝振纯,王国庆,等.新安江模型和改进BP神经网络模型在闽江水文预报中的应用[J].水资源与水工程学报,2017,28(1):40-44.
- [18] 王庆杰,岳春芳,李艺珍.基于MAGA-PPC模型的水资源配置方案综合评价[J].水资源与水工程学报,2018,29(3):105-110.

(上接第132页)

- [6] 何鑫,吴吉东,李颖,等.基于SPEI的辽西地区气象干旱时空分布特征[J].干旱区地理,2017,40(2):340-347.
- [7] 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.
- [8] 郭树龙,温季,姜新.基于SPEI的1960-2015年江汉平原旱涝规律分析及预测[J].灌溉排水学报,2018,37(9):108-115.
- [9] 蔡思扬,左德鹏,徐宗学,等.基于SPEI干旱指数的东北地区干旱时空分布特征[J].南水北调与水利科技,2017,15(5):15-21.
- [10] 熊光洁,张博凯,李崇银,等.基于SPEI的中国西南地区1961-2012年干旱变化特征分析[J].气候变化研究进展,2013,9(3):192-198.
- [11] 任启伟,李鑫华,尹小玲,等.广东省干旱灾害识别及变化趋势[J].热带地理,2017,37(4):494-500.
- [12] 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):382-390.
- [13] 徐晗.陕西省干旱时空演变特征分析及脆弱性评价[D].西安:长安大学,2018.
- [14] 郭旭新,赵英,高志永,等.基于SPEI的陕北黄土丘陵区干旱特征及影响因素分析[J].西北林学院学报,2019,34(1):69-76.
- [15] 轩俊伟,郑江华,刘志辉.基于SPEI的新疆干旱时空变化特征[J].干旱区研究,2016,33(2):338-344.
- [16] 赵静.黄淮海流域天然植被对干旱的响应及耐旱特征[D].上海:东华大学,2015.
- [17] 庄少伟,左洪超,任鹏程,等.标准化降水蒸发指数在中国区域的应用[J].气候与环境研究,2013,18(5):617-625.
- [18] 李英杰.陕西省旱涝灾害多时空尺度特征与趋势判断[D].西安:陕西师范大学,2017.
- [19] 陈学凯.贵州省多时间尺度气象干旱时空变化特征研究[D].郑州:华北水利水电大学,2016.
- [20] 张钰,孙涛,方国华,等.新乡市城区近64年气温和降水量特征分析[J].水资源与水工程学报,2017,28(1):94-99.
- [21] 夏伟,周维博,李文溢,等.气候变化和人类活动对泮河流域径流量影响的定量评估[J].水资源与水工程学报,2018,29(6):47-52.