

多元线性回归与BP神经网络模型在次洪量预测中的对比研究

冯鑫伟, 黄领梅, 沈冰

(西安理工大学 西北旱区生态水利工程国家重点实验室培育基地, 陕西 西安 710048)

摘要: 针对半干旱地区次洪量预测问题, 选取岔巴沟流域曹坪水文站 1980-2010 年中 15 场洪水资料, 根据实测次暴雨、洪量资料, 考虑淤地坝控制面积、次暴雨量、暴雨中心位置、前期影响雨量等因子, 利用 SPSS 及 MATLAB 软件, 建立用以预测次洪量的多元线性回归模型和 BP 神经网络模型。模型预测结果比较表明: 多元线性回归模型和 BP 神经网络模型都能较好地应用于次洪量的预测, 进一步得出 BP 神经网络模型的预测效果优于多元线性回归模型。研究结果可为淤地坝的安全度汛提供决策依据。

关键词: 淤地坝; 次洪量预报; 多元线性回归; BP 神经网络;

中图分类号: TV124

文献标识码: A

文章编号: 1672-643X(2017)03-0123-04

Comparative study on multivariate linear regression and BP neural network model in the prediction of flood volume

FENG Xinwei, HUANG Lingmei, SHEN Bing

(State Key Lab Cultivation Base of Northwest Arid Ecology and Hydraulic Engineering,
Xi'an University of Technology, Xi'an 710048, China)

Abstract: Aiming at the problem of flood volume forecast in semi-arid area, we selected 15 flood data of the Caoping hydrological station in Chabagou watershed from 1980 to 2010. According to the measured single storm, flood control data in Chabagou watershed, considering the warp land dam area, storm rainfall, rainfall center, antecedent rainfall and others factors, combined with SPSS and MATLAB software, multivariate linear regression model and BP neural network model for the prediction of flood volume were built. The prediction results of the two models show that multivariate linear regression model and BP neural network model can be better applied to flood volume prediction, and the BP neural network model is superior to multivariate linear regression model. Research result can provide decision-making basis for dam safety in the flood season.

Key words: warp land dam; flood volume prediction; multiple linear regression; BP neural network

淤地坝是构筑于水土流失地区各级沟道中, 用以稳定和抬高侵蚀基点、拦泥淤地、滞洪削峰的坝工建筑物, 是黄土丘陵沟壑区重要的水土保持工程措施之一, 在拦蓄泥沙、防洪保收、改善交通、农业生产等方面发挥着重要作用^[1-4]。

新中国成立后, 淤地坝建设迅速发展, 但同时淤地坝水毁事件也频频发生, 遇到暴雨, 尤其是特大暴

雨经常会发生水毁垮坝事件, 多数甚至是连锁垮坝。尤其是 20 世纪 70 年代, 黄河中游地区连续几次大暴雨, 对该地区的淤地坝造成不同程度的损坏, 给当地人民群众的生命财产安全造成严重威胁, 以至于学术界对于淤地坝的建设有了异议, 甚至是偏见^[5-7]。因此多年的水毁问题一直困扰着黄土高原地区淤地坝的建设发展。

收稿日期: 2016-12-31; 修回日期: 2017-02-21

基金项目: 国家自然科学基金项目(51679184); 陕西省水利厅项目(2016slkj-12)

作者简介: 冯鑫伟(1991-), 男, 陕西榆林人, 硕士研究生, 主要从事水文学及水资源研究。

通讯作者: 黄领梅(1972-), 女, 四川乐至人, 博士, 副教授, 主要从事水文学及水资源研究。

综上所述,虽然在建设淤地坝时已经充分考虑了洪水的问题,但是超标准洪水仍可能发生,此外由于泥沙淤积导致坝库蓄洪、滞洪能力降低也是暴雨洪水引起溃坝的直接原因^[6]。因此精准的次洪量预测预警已经成为淤地坝安全度汛的重要课题^[8-11]。

由于次洪量预报的影响因素较多,而这些因素大部分是随机的或未知的,这使得预报的精度和预见期受到了很大的制约。人工神经网络^[12-14]因其独特的非线性、非凸性、自适应性和处理各种信息的能力,被证明可应用到洪水预报中。多元线性回归^[15]利用线性来拟合多个自变量和因变量之间的关系,从而确定多元线性回归的参数,进而回归至原假设方程中预测因变量的趋势。因此,本文将分别建立BP神经网络模型和多元线性回归模型,对两种模型的预测效果进行对比研究。

1 研究区概况及数据来源

岔巴沟流域位于东经 109°47', 北纬 37°31' 之间,集水面积 205 km²,沟道长 26.3 km。岔巴沟隶属黄河水系,位于黄土丘陵沟壑区第一副区,在无定河流域的西南部和无定河的支流大理河相汇。流域出口处设有曹坪水文站,控制面积 187 km²。岔巴沟流域属于干燥少雨的大陆性气候,多年平均降雨量 450 mm 左右,年内分配不均匀。降雨主要集中在汛期(6-9月),占年雨量的 70%。该流域内暴雨洪水的特点为暴雨历时短、雨强大、次洪量大,全年降雨量主要集中于几场暴雨,每逢暴雨干支流均会出现较大的洪水。

岔巴沟流域分布有骨干坝 135 座,其中骨干坝 23 座,中型坝 112 座。流域内淤地坝工程建设数量大且分布较均匀。在仅考虑骨干坝的控制流域面积下,岔巴沟流域被控制面积占该流域总面积的 41.1%,这些淤地坝工程改变了流域下垫面情况,直接对产汇流造成影响^[7]。

本次研究选取岔巴沟流域曹坪水文站 1980-2010 年的 15 场洪水资料以及流域范围内雨量站所观测和记载的降雨资料,其中 12 场用于建立模型,其余 3 场洪水用于检验模型。降雨资料选自岔巴沟流域内的姬家岑、曹坪、朱家阳村、和民焉、杜家山、牛薛沟、桃园山、万家堰、刘家抓、马虎焉,共 10 个雨量站;洪水资料采用岔巴沟流域沟口的曹坪水文站。岔巴沟流域位置与站点分布见图 1。

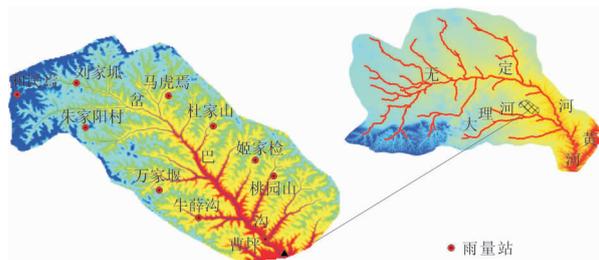


图1 岔巴沟流域位置及水文站分布图

2 研究方法

2.1 多元线性回归模型

多元线性回归模型^[8]是用 2 个或 2 个以上的解释变量来解释因变量的一种计量模型。设 Y 为因变量, X_1, X_2, \dots, X_n 为 n 个用来说明 Y 的被称为解释变量的不同变量,则:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon \quad (1)$$

式中: β_0 是常数项,又被称作截距; $\beta_1, \beta_2, \dots, \beta_m$ 为回归系数,表示在其他自变量均保持不变时, X_i ($i = 1, 2, \dots, m$) 每改变一个单位,因变量 Y 的平均变化量; ε 是去除 m 个自变量对 Y 影响后的随机误差,又称残差。

多元线性回归模型建成之后,必须进行评价和检验后才能判定模型是否可以应用。模型的检验主要通过考察模型的拟合优度、显著性检验来综合评价模型的可应用性。

2.2 BP神经网络模型

BP神经网络是目前应用最广泛的 ANN 模型之一。它是一种基于误差反向传播算法训练的多层感知器前馈网络,具有以非线性适应性为特征的信息处理能力、十分优良的非线性逼近能力以及较好的容错性,因而能够对大规模数据进行并行处理,自组织、自学习与自适应能力突出,被广泛应用于预测、分类、模式识别和聚类等领域。另外这种反向传播算法使得多层感知器具备逼近任意复杂的非线性函数的能力。BP神经网络结构如图 2 所示。

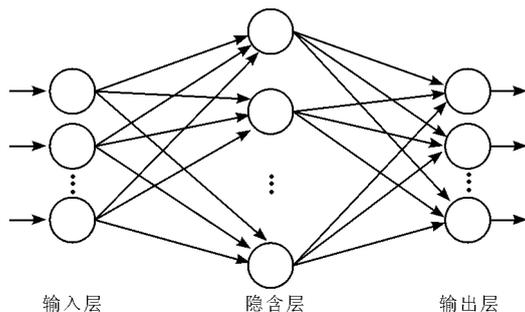


图2 三层BP神经网络结构图

3 结果与分析

3.1 次洪量的影响因素

影响次洪量大小的因素分为天气因素和流域下垫面因素两类。其中天气因素是影响次洪量的主要因素,黄土高原的大洪水主要由暴雨形成。次暴雨的总量、暴雨强度及其时空分布,决定了流域洪水大小和特征。流域下垫面因素主要是指土地利用变化、植被因素以及淤地坝数量等。根据以上分析,选取 1986、1995 以及 2008 年黄土高原 250 × 250 m 土地利用资料,从中提取出岔巴沟流域的土地利用数据(见表 1)。

表 1 岔巴沟土地利用情况统计表

类型	1986 年		1995 年		2008 年	
	面积/ km ²	比例/ %	面积/ km ²	比例/ %	面积/ km ²	比例/ %
林地	7.0	3.8	4.8	2.6	7.0	3.7
草地	73.6	39.3	75.1	40.1	72.9	39.0
耕地	106.2	56.8	106.5	57.0	106.7	57.1
其它	0.2	0.1	0.5	0.3	0.4	0.2
总计面积	187.0	100.0	187.0	100.0	187.0	100.0

结果表明,岔巴沟流域在 20 多年内土地利用变化较小,对次洪量大小的影响可忽略不计,因此次洪量的影响因素可不考虑土地利用变化。

根据上述综合分析,最终确定淤地坝控制面积(km²)、次暴雨量(mm)、暴雨中心位置(m)、前期影

响雨量(mm)、洪峰流量(m³/s)^[16]以及最大 3 h 雨强(mm/min)6 项指标作为影响次洪量的因素,其中暴雨中心采用次降雨最大的雨量站到出口的距离。

3.2 多元线性回归模型

本次研究根据 SPSS 软件进行计算建模,根据上述分析,选取淤地坝控制面积(X_1)、次暴雨量(X_2)、暴雨中心位置(X_3)、前期影响雨量(X_4)、洪峰流量(X_5)以及最大 3 h 雨强(X_6)作为自变量,次洪量为因变量,对变量进行无量纲化。由于影响因素众多,并且不能判定各影响因素对次洪量的影响是否显著,为建立最优回归模型,采用逐步回归方法。运用 SPSS 软件的多元回归分析工具进行逐步回归拟合,可得 6 个影响因素同次洪量相关性都较强,无可移除变量。

根据表 2 显示, $R = 0.969$, 接近于 1, Durbin - Watson 检验统计量为 2.752, 说明残差独立, 由此说明方程的拟合优度较高, 自变量可以解释因变量的变化。

表 2 多元线性回归模型汇总表

模型	R	R ²	标准估计的误差	Durbin - Watson
1	0.969	0.939	25.995	2.752

根据 SPSS 分析表显示, 显著性检验 P 值为 0.007。本次取 95% 的置信区间($\alpha = 0.05$), $P = 0.007 < \alpha = 0.05$, 表示回归模型整体解释变异量达到显著性水平, 说明自变量和因变量之间存在显著线性关系。

表 3 多元线性回归系数表

变量符号	非标准化系数		标准系数 试用版	t	Sig.	共线性统计量	
	B	标准误差				容差	VIF
常量	-71.716	50.138		-1.143	0.212		
X_1	0.372	0.337	0.166	1.105	0.319	0.536	1.867
X_2	1.375	0.306	0.686	4.496	0.006	0.519	1.925
X_3	-2.641	1.878	-0.174	-1.407	0.219	0.793	1.260
X_4	0.996	0.683	0.226	1.459	0.204	0.505	1.978
X_5	0.209	0.124	0.279	1.684	0.153	0.443	2.260
X_6	235.738	170.138	0.202	1.386	0.225	0.570	1.755

根据模型处理结果获得回归系数, 通过共线性检验结果得出模型不存在共线性问题。由非标准化系数获得拟合结果:

$$Y = -71.716 + 0.372X_1 + 1.375X_2 - 2.641X_3 +$$

$$0.996X_4 + 0.209X_5 + 235.738X_6$$

3.3 BP 神经网络模型

本次研究同样选取 3.2 中 6 个影响因素作为 BP 神经网络的输入层, 将次洪量作为网络的输出

层。设置最大迭代次数为 10000 次,网络目标误差为 0.000001,学习速度为 0.01。经过经验公式和反复对比分析,比较输出误差的大小,最终确定该神经网络隐层神经元单位数目为 20。

根据以上设置进行 BP 神经网络的训练,直至符合预定的目标,经测试神经网络在 43 步时收敛速度较快且达到要求,总体 $R = 0.96986$, Grident 为 0.443, validation Checks 为 6,模型具有较好的拟合优度。

3.4 模型比较研究

3.4.1 建模比较 利用 12 次暴雨洪量资料建立的多元线性回归模型与 BP 神经网络模型,模拟值与实测值之间的拟合度 (R^2) 分别是 0.939 与 0.941。相比较而言,BP 神经网络模型具有较高的拟合优度及良好的建模效果。

12 次暴雨洪量资料对应的模拟误差见表 4 和图 3。由表 4 可以看出,BP 神经网络模型与多元线性回归模型的平均相对误差分别是 18.97% 与 8.66%,因此 BP 神经网络模型在建模期的效果优于多元线性回归模型。由图 3 可知,两种模型对样本拟合度均比较高 ($R^2 > 0.9$),但是 BP 神经网络模型的精度高于多元线性回归模型,能达到更好的拟合效果。

表 4 两种模型在建模期内的相对误差

洪号	实际值/ 10^4 m^3	多元线性回归模型		BP 神经网络模型	
		模拟值/ 10^4 m^3	相对 误差/%	模拟值/ 10^4 m^3	相对 误差/%
1	99.33	91.56	7.8	92.73	6.6
2	163.8	152.38	7.0	139.2	15.0
3	178.97	184.23	-2.9	183.68	-2.6
4	35.89	52.67	-46.8	42.21	-17.6
5	52.69	59.20	-12.4	46.75	11.3
6	45.07	35.19	21.9	51.61	-14.5
7	128.70	114.56	11.0	138.4	-7.5
8	155.90	171.01	-9.7	150.8	3.3
9	249.57	255.71	-2.5	247.32	0.9
10	74.43	52.77	29.1	82.95	-11.4
11	50.70	87.12	-71.9	52.5	-3.6
12	207.93	186.08	10.5	188.03	9.6

3.4.2 预测能力比较 为进一步检验多元线性回归模型和 BP 神经网络模型的预测能力,将未参与建模的 3 场暴雨洪水资料,分别输入模型得到结果,

见表 5。

表 5 检验期的预测值与实测值对比

洪号	实测值/ 10^4 m^3	线性回归模型		BP 神经网络模型	
		预测值/ 10^4 m^3	相对 误差/%	预测值/ 10^4 m^3	相对 误差/%
19880713	64.41	69.77	-8.3	72.70	-3.9
19950713	50.50	60.48	-19.8	51.47	-1.9
20040819	65.56	55.33	15.6	71.44	-9.0

从表 5 可以看出,回归模型预测偏差的平均相对误差为 14.6%,BP 神经网络模型的预测平均相对误差为 4.9%。由此可见两种模型对新样本的预测能力都能满足要求,但 BP 神经网络模型比线性回归模型更加精确。

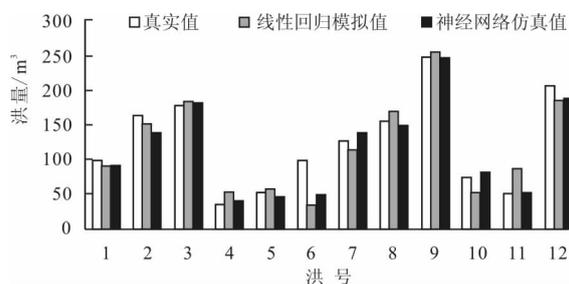


图 3 两种模型的模拟值与实测值柱状图

由以上分析知,BP 神经网络模型对于次洪量的预测较之多元线性回归的预测更为敏感和精确,预测值的精度高于多元线性回归模型。研究分析这种现象的原因是由两种模型自身的特点、次洪量与各影响因素的关系以及各影响因素内部之间相关关系共同决定的。

由于因变量次洪量与其影响因素之间并非完全典型的线性相关关系,同时次洪量的各个影响因素之间具有一定的相关关系,多元线性回归模型无法准确地描述这种非线性相关关系,因而会造成一定的偏差。而 BP 神经网络模型属于隐含式的数学方法,在分析过程中不需要建立数学模型,只需要通过原始数据训练得到稳定的学习网络,进而对数据进行预测,这种方法更加简便快捷。

因此,神经网络模型可以更好地模拟非线性方程,次洪量与其影响因素的非完全线性相关关系以及各个影响因素内部之间的相关关系不会对神经网络的预测产生影响,进而在一定情况下提高了神经网络的预测精度。

(下转第 133 页)

参考文献:

- [1] 刘敏,权瑞松,许世远.城市暴雨内涝灾害风险评估[M].北京:科学出版社,2012.
- [2] 初祁,彭定志,徐宗学,等.基于 MIKE 11 和 MIKE 21 的城市暴雨洪涝灾害风险分析[J].北京师范大学学报(自然科学版),2014,50(5):446-451.
- [3] Renö, Sellke P. Risk, society and policy making: Risk governance in a complex world[J]. International Journal of Performability Engineering, 2011,7(4):349-366.
- [4] 赵燕霞.基于水力学模型的城市除涝方案分析研究[D].大连:大连理工大学,2014.
- [5] 郭凤清,屈寒飞,曾辉,等.基于 MIKE21 的滙江蓄滞洪区洪水危险性快速预测[J].自然灾害学报,2013,22(3):144-152.
- [6] 丛翔宇,倪广恒,惠士博,等.基于 SWMM 的北京市典型城区暴雨洪水模拟分析[J].水利水电技术,2006,37(4):64-67.
- [7] 黄卡. SMS 模型在长洲水利枢纽下游水流模拟中的应用研究[J].红水河,2010,29(3):24-28.

- [8] 叶爱民,刘曙光,韩超,等. MIKE FLOOD 耦合模型在杭嘉湖流域嘉兴地区洪水风险图编制工作中的应用[J].中国防汛抗旱,2016,26(2):56-60.
- [9] 陈虹,雷少青,王胜安.福州市大学新区暴雨洪水积涝数学模型研究与应用[J].福州大学学报(自然科学版),2011,39(1):137-142.
- [10] 邵尧明,邵丹娜.城市暴雨积水过程的计算方法[J].中国给水排水,2008,24(6):53-55.
- [11] 刘绍青.济南市城区洪水淹没模拟研究[D].济南:山东大学,2009.
- [12] 黄程. MIKE 11 模型在东莞市防洪分析中的应用[J].水利规划与设计,2013(4):26-29.
- [13] 梁彬锐. MIKE11 模型在沙井河片区防洪排涝工程中的应用[J].中国农村水利水电,2008(7):81-83.
- [14] 姬战生,张飞珍,孙映宏.杭州市上塘河流域防洪能力分析[J].浙江水利科技,2012(6):46-48.
- [15] 朱婷,王鑫.基于 MIKE FLOOD 模型的中顺大围洪水风险研究[J].中国水运,2016,37(7):71-74.
- [16] 曹永强,黄林显,苗迪,等.我国洪水风险图绘制方法研究[J].人民黄河,2008,30(8):6-7.

(上接第 126 页)

4 结 论

(1) 本文综合分析了岔巴沟流域内影响次洪量的主要因素,选取淤地坝控制面积、次暴雨量、暴雨中心位置、前期影响雨量、洪峰流量以及最大 3 h 雨强 6 项指标作为影响次洪量的指标因素,建立了预测次洪量的 BP 神经网络模型和多元线性回归模型,结果表明两种模型均能很好地应用于次洪量的预测。

(2) 通过对比分析,BP 神经网络模型的模拟与预测效果均好于多元线性回归模型;BP 神经网络的非线性映射关系对变量的处理效果优于线性回归模型的线性关系。神经网络在次洪量的预测中将得到更广泛的应用,可为淤地坝的防洪预警提供决策依据。

(3) 未来可以将多元线性回归模型与 BP 神经网络模型进行组合,即将多元线性回归的线性关系和 BP 神经网络的非线性映射关系充分考虑在次洪量的预测中,以提高预测精度。

参考文献:

- [1] 高海东,贾莲莲,李占斌,等.基于图论的淤地坝对径流影响的机制[J].中国水土保持科学,2015,13(4):1-8.
- [2] 许灏.淤地坝——黄土高原群众致富增收的关键[J].水利发展研究,2005,5(5):39-41.
- [3] 黄河中上游管理局.黄河流域水土保持概论[M].郑州:黄河水利出版社,2011:88-92.
- [4] 杨筱筱.水土保持措施对于秃尾河流域产汇流参数的影

- 响研究[D].杨凌:西北农林科技大学,2012.
- [5] 许五弟,杨勤科,梁剑辉.淤地坝溃坝预报预警地理信息模型初探[J].中国水土保持 SWCC,2010,14(1):42-43.
- [6] 魏霞,李占斌,武金慧.淤地坝水毁灾害研究中的几个观念问题讨论[J].水土保持研究,2007,14(6):235-237.
- [7] 李莉,王峰,孙维营,等.黄土高原淤地坝水毁问题分析[J].中国水土保持,2014,18(10):20-22.
- [8] 李靖,张金柱,王晓.20世纪70年代淤地坝水毁灾害原因分析[J].中国水利,2003(17):55-56.
- [9] 肖洁.多模型洪水预报方法研究及系统实现[D].西安:西安理工大学,2013.
- [10] 权琦泽.库坝影响下黄土地区设计洪峰研究[D].西安:西安理工大学,2016.
- [11] 韩瑞光,冯平.流域下垫面变化对洪水径流影响的研究[J].干旱区资源与环境,2010,24(8):27-30.
- [12] 张景阳,潘光友.多元线性回归与 BP 神经网络预测模型对比与运用研究[J].昆明理工大学报(自然科学版),2013,38(6):61-67.
- [13] 丁华,李炎蔚,袁冬. BP 神经网络预测激光微孔表面粗糙度[J].激光与光电子学进展,2017,54(1):1-13.
- [14] 赵亮,刘吉平,徐艳艳.基于 BP 神经网络模型的三江平原湿地面积预测研究[J].干旱区资源与环境,2012,26(10):53-56.
- [15] 王素立,刘永.基于波动相关性及主分量变换的多元线性回归模型研究[J].统计与决策,2012(22):18-21.
- [16] 周玲微,雷廷武,武阳.岔巴沟流域次暴雨产流无量纲模型[J].农业工程学报,2010,26(11):54-60.