

# 基于神经网络及灰色关联分析的土壤入渗研究\*

黄俊<sup>1,2</sup> 吴普特<sup>1,2,3†</sup> 赵西宁<sup>1,2,3</sup>

(1 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100)

(2 国家节水灌溉杨凌工程技术研究中心, 陕西杨凌 712100)

(3 中国科学院水利部水土保持研究所, 陕西杨凌 712100)

## RESEARCH ON SOIL INFILTRATION BASED ON BP NETWORK AND GRAY CORRELATION ANALYSIS

Huang Jun<sup>1,2</sup> Wu Pute<sup>1,2,3†</sup> Zhao Xining<sup>1,2,3</sup>

(1 College of Water Resources and Architecture Engineering, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

(2 National Engineering Research Center for Water Saving Irrigation at Yangling, Yangling, Shaanxi 712100, China)

(3 Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China)

关键词 土壤入渗; BP神经网络; 缺省因子检验; 灰色关联分析

中图分类号 S152.7 文献标识码 A

入渗系水分自土表垂直向下进入土壤的过程, 也包括侧向及向上进入土壤的过程<sup>[1]</sup>, 它是自然界水循环中的一个重要环节。土壤入渗研究对于减少地表径流、降低土壤侵蚀和搞好生态环境建设等方面具有重要的理论及现实意义, 国内外许多学者投入了大量的研究工作, 并取得了诸多重要的研究成果。达西定律开创了入渗研究的土壤物理学途径, Richards 于 1931 年在达西定律的基础上导出非饱和流方程<sup>[2]</sup>。Green 和 Ampt 以及 Smith 等分别建立了特定条件下土壤入渗的物理模型, 其中 Green 和 Ampt<sup>[3]</sup> 及 Philip<sup>[4]</sup> 模型均是基于积水条件, 这显然与实际降雨并不相同。随后, Smith<sup>[5]</sup> 及 Smith 和 Parlange<sup>[6]</sup> 等提出了降雨条件下、可计算地表积水时间的入渗模型。同时, 经验的入渗公式由于参数简单、使用方便也得到了广泛的应用, 如 Kostiaiko 公式<sup>[7]</sup>、Horton 公式<sup>[8]</sup> 等。我国学者<sup>[9-13]</sup> 也在土壤入渗的研究方面做了大量的工作。入渗经验公式的参数简单但无实际物理意义, 通过对入渗非线性过

程的简化而对入渗规律进行简单定量的描述; 而物理模型虽能较为准确的描述入渗的非线性过程, 但其参数复杂求解麻烦。影响土壤入渗的因子较多, 如地形地貌、土壤性质、地表植被等。关于单因子土壤入渗的研究较多, 但缺少各因子与土壤入渗关联度以及各因子敏感性的相关分析。本文选取降雨强度、土壤前期含水量、植被覆盖度和坡度 4 因子作为影响坡面土壤入渗的主要因子, 建立坡面土壤稳定入渗率 BP 模型, 运用缺省因子检验法分析了各因子对土壤稳定入渗率的敏感性, 并采用灰色关联分析法对各因子进行综合排序, 以验证 BP 缺省因子检验法的正确性, 同时也完成了灰色关联分析的自我检验。

## 1 材料与方法

试验在陕西省杨陵区五泉乡岭后中国科学院水土保持研究所坡面降雨径流调控利用野外定位

\* 国家自然科学基金项目(40701092)、国家 863 计划项目(2006AA100217) 及国家科技支撑计划项目(2007BAD88B10; 2007BAD88B05) 资助

† 通讯作者, E-mail: gjzwpt@vip.sina.com

作者简介: 黄俊(1983—), 男, 河南新蔡县人, 博士研究生, 从事农业水土资源高效利用研究。E-mail: hold-onyourdream@yahoo.com.cn

收稿日期: 2010-01-26; 收到修改稿日期: 2010-11-20

监测站进行。采用自制的针头式人工模拟降雨装置,雨强范围为  $0.3 \sim 3 \text{ mm min}^{-1}$ ,降雨高度为  $1 \sim 1.7 \text{ m}$ ,最大有效降雨面积  $1 \text{ m} \times 1.5 \text{ m}$ ,降雨器周围有遮风帘,可减少风对人工降雨试验的不利影响,参考黄土高原坡面典型降雨特征,选定的降雨强度为  $0.5、0.75、1.0、1.5$  和  $2.0 \text{ mm min}^{-1}$  五个等级。供试土样采自杨凌示范区节水博览园与五泉乡岭后,过  $10 \text{ mm}$  筛后按质量比  $1:1$  配制而成,土壤类型为砂粉土(按我国土壤质地分类标准),土壤机械组成采用 MS2000 激光粒度仪(英国马尔文仪器有限公司生产)测定(表 1)。

表 1 供试土壤的机械组成

粒径范围(mm)	含量(%)
0.25 ~ 0.05	13.27
0.05 ~ 0.01	42.82
0.01 ~ 0.005	12.44
0.005 ~ 0.001	19.00
< 0.001	0.31

受雨坡面是可变坡度的移动式装土槽,长  $\times$  宽  $\times$  高 =  $1.2 \text{ m} \times 0.8 \text{ m} \times 0.45 \text{ m}$ ,下端有导流槽用于坡面产流后浑水的排出。装土槽底部有排水孔,填土前,先在槽底铺上一层纱布,便于土壤水分均匀下渗。土壤干容重控制在  $1.35 \sim 1.40 \text{ g cm}^{-3}$ ,含水量约为  $6\%$ ,从土槽底部起,每隔  $10 \text{ cm}$  填装一层,共 4 层,深  $40 \text{ cm}$ ,边装边均匀压实,每次在填装下一层前将前层土打毛,以消除两层土壤间层理对入渗的不利影响。本试验中生物调控材料有两种:牧草和农作物,分别为黑麦草(*Lolium perenne* L)、苜蓿(*Medicago sativa* Linn)和春小麦(*Triticum aestivum*)。植被覆盖度测定方法:用数码相机采集(相机镜头与坡面垂直,且垂直距离基本不变)坡面植被生长状况( $3 \sim 5$  张照片,为 jpg 格式),然后采用 Photoshop 和 ImageJ 软件处理得到植被的覆盖度( $\%$ ),并结合目估法校正。

试验中蒸散、填注和植物截留水量很少,忽略不计。故任一时刻土壤入渗率为:

$$i = i_r \times \cos\theta - 10 \times r / (s \times t) \quad (1)$$

式中, $r$  为径流量( $\text{ml}$ ), $s$  为坡面受雨面积( $\text{cm}^2$ ), $t$  为降雨历时( $\text{min}$ ), $i$  为土壤入渗率( $\text{mm min}^{-1}$ ), $i_r$  为降雨强度( $\text{mm min}^{-1}$ ), $\theta$  为坡度( $^\circ$ )。

土壤稳定入渗率由入渗历时曲线通过 Horton 公式<sup>[2]</sup>拟合得到:

$$i = i_0 + (i_c - i_0) e^{-fp^t} \quad (2)$$

式中, $i_0$  为土壤稳定入渗率( $\text{mm min}^{-1}$ ), $i_c$  为土壤初始入渗率( $\text{mm min}^{-1}$ ), $fp$  为拟合参数, $i$  为土壤入渗率( $\text{mm min}^{-1}$ ), $t$  为降雨历时( $\text{min}$ ), $e$  为自然对数底。本文筛选了 51 组人工模拟降雨实测数据,对其稳定入渗率进行拟合,各拟合方程的决定系数均大于  $0.9$ ,表明拟合得到的数据真实可信。

## 2 坡面土壤稳定入渗率 BP 模型

### 2.1 BP 神经网络及其算法改进

人工神经网络(ANN)系统是 20 世纪 40 年代后出现的,它是由众多神经元经权值相互连接组成,具有大规模并行处理、分布式信息存储、良好的自组织自学习能力,在信息处理、模式识别、智能控制及系统建模等领域得到越来越广泛的应用<sup>[14]</sup>。ANN 中的 BP 网络(Error Back Propagation-前反馈神经网络,简称为 BP)尤为突出,它可以逼近任意连续函数,具有很强的非线性映射能力,且网络的中间层数、各层的神经元数及学习率等可适当调整,灵活性较大,因此 BP 网络得到了广泛应用。其算法的基本思想<sup>[15]</sup>:学习信息向前传播,通过误差的反向传播而不断调整网络权值和偏差进行反复训练,使输出值与期望值尽可能接近,当网络输出误差小于指定误差或达到最大迭代次数时停止训练。

BP 算法的主要缺点是训练速度较慢,本文采用附加动量法和学习速率自适应调整两种策略对其进行改进<sup>[16-17]</sup>。动量项反映了网络以前积累的调整经验,对于  $t$  时刻网络调整起到了阻尼作用,当误差曲面出现骤然起伏时,可有效抑制网络陷入局部极值状态。若  $W$  表示某层权矩阵, $X$  表示某层输入向量,则含有动量项的权值调整向量表达式为:

$$\Delta W(tn) = \eta \delta X + \alpha \Delta W(tn - 1) \quad (3)$$

式中, $\alpha$  为动量系数,一般有  $0 < \alpha < 1$ ,文中取  $0.95$ <sup>[18]</sup>;  $\delta$  为误差曲面斜率; $\eta$  为学习率; $tn$  为训练次数。

标准 BP 算法中的学习率是固定的,这可能会因为其过大而跳过较窄的“坑凹”而增加网络的迭代次数。而自适应学习速率是通过网络误差下降快慢来进行自我调整,从而保证网络以较好的学习速率进行训练,能增加网络收敛速度、缩短网络学习时间。自适应学习率表达式为:

$$\eta(k+1) = \begin{cases} \beta\eta(k) & E_{RME}(k+1) < E_{RME}(k) \\ \theta\eta(k) & E_{RME}(k+1) > 1.04E_{RME}(k) \\ \eta(k) & \text{其他} \end{cases} \quad (4)$$

式中,  $\beta$ 、 $\theta$  均为学习速率调整比例, 且  $\beta > 1$ ,  $\theta < 1$ , 一般取  $\beta = 1.05$ ,  $\theta = 0.7$ <sup>[19]</sup>;  $E_{RME}$  为网络总误差。

## 2.2 土壤稳定入渗率 BP 模型的建立以及与多元回归模型的对比分析

研究表明, 3 层 BP 网络能够逼近或映射多数有理函数<sup>[20]</sup>, 故本文选取 3 层 BP 网络对土壤入渗进行模拟, 包括输入层、隐含层及输出层。隐含层采用双曲正切 S 型函数  $f(x) = 1/(1 + e^{-x})$  为 BP 网络的传递函数, 而输出层采用线性函数。为减少网络训练过程中权值调整幅度和网络训练时间, 用式 (5) 对样本进行归一化处理<sup>[21]</sup>。

$$X_i = x_i / (\min(x_i) + \max(x_i)) \quad (5)$$

式中,  $X_i$  为归一化后第  $i$  个样本输入值,  $x_i$  为实际样本第  $i$  个输入值;  $\min(x_i)$  和  $\max(x_i)$  分别为样本序列最小值和最大值。

运用 Matlab 神经网络工具箱编写程序, 建立 4 个输入因子的土壤稳定入渗率 BP 模型, 并进行训练。通过多次对隐含层神经元结点数的调整及试算, 认为 10 个隐含层神经元节点为最优, 网络训练中收敛速度及误差控制能力较好, 同时得到拟合及检验效果也较为理想, 故网络拓扑结构为 4:10:1。

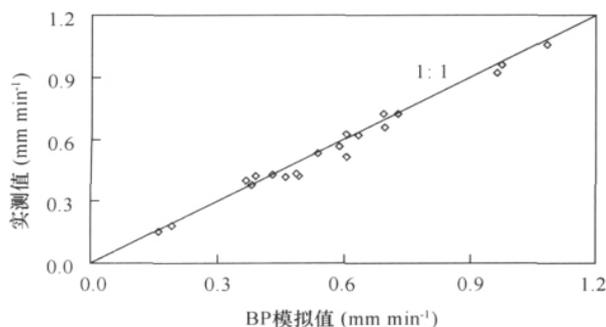


图1 土壤稳定入渗率 BP 模型检验

文中采用 51 组实测数据中的 31 组作为训练样本, 另外的 20 组作为检验样本。网络训练总误差为  $0.0024 \text{ mm min}^{-1}$ , 实测值与模拟值间回归方程为:  $Y_{\text{实测}} = 0.9710 X_{\text{BP模拟}}$ ,  $R^2 = 0.9820$ , 且达到极显著水平 ( $F = 970.7 > F_{0.01}(1, 18) = 8.29$ )。说明 BP 模型模拟结果比较理想, 改进后的 BP 神经网络能准确模拟不同条件下土壤入渗状况。

对 31 组实测数据进行多元回归分析, 得到土壤稳定入渗率多元回归模型:

$$\ln(i_0) = -0.957 + 0.469 \ln(i_r) - 0.241 \ln(w) - 0.230 \ln(S) + 0.860 \ln(vc) \quad (R^2 = 0.786, \alpha = 0.05) \quad (6)$$

式中,  $i_r$  为降雨强度 ( $\text{mm min}^{-1}$ );  $w$  为土壤前期含

水量 (%);  $S$  为坡度 ( $^\circ$ );  $vc$  植被覆盖度 (%).

由图 2 可以看出, BP 模拟值的相对误差波动小、较为集中, 计算结果更为稳定。其计算结果仅有 4 组样本的相对误差大于 10%, 而多元回归模型有近 50% 的相对误差超过 10%。20 组样本的 BP 模型与多元回归模型计算值相对误差平均值和变异系数分别为: 5.6% 和 0.9、11.8% 和 1.2, 说明 BP 模型较回归模型有更好的稳定性和非线性映射能力, 能更为准确地定量化土壤入渗过程。

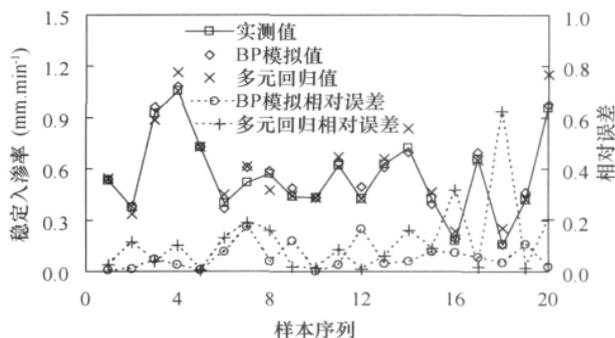


图2 土壤稳定入渗率 BP 模拟值、多元回归值和实测值比较

## 3 基于 BP 模型的土壤稳定入渗率敏感性分析

### 3.1 缺省因子检验法

上述 BP 模型的建立考虑了多因子对土壤入渗的影响, 但在实际操作中许多因子很难同时把握, 这就需要掌握影响土壤稳定入渗率最为强烈的因子。本文针对各影响因子进行人工神经网络模型缺省因子检验, 即通过对各输入因子的缺省进行神经网络模型的建立及检验, 根据其检验误差与全因子模型检验误差的比值  $R_i$  大小来确定缺省因子对输出因子响应的敏感程度。

$$R_i = RMSE_i / RMSE \quad (7)$$

式中,  $R_i$  为第  $i$  个因子的敏感性指数;  $RMSE_i$  为缺省第  $i$  个因子时模型检验误差;  $RMSE$  为全因子模型检验误差。若  $RMSE_i > RMSE_j$ , 则表示土壤稳定入渗率对第  $i$  个因子较第  $j$  个因子敏感。若  $R_i > 1$ , 表明该因子为土壤稳定入渗率的敏感因子。

### 3.2 敏感性分析结果

敏感性分析时, 采用全部 51 组实测数据作为样本, 然后逐一进行输入层因子缺省检验, 采用 2.1 节中改进的 BP 算法, 建立 3 因子土壤稳定入渗率模型。由表 2 可知, 4 个因子对土壤稳定入渗率的敏感程度依次为: 降雨强度 > 植被覆盖度 > 土壤前期

含水量 > 坡度。降雨强度敏感性指数为 2.546, 是 4 因子中最大值, 说明降雨强度是影响土壤入渗的主导因子。植被覆盖度敏感性指数为 1.536, 是影响土壤入渗的重要因子。而土壤前期含水量的敏感

性指数仅为 1.164, 是影响土壤稳定入渗率的敏感因子。坡度敏感指数接近于 1, 说明坡度对土壤稳定入渗率的影响作用不大。

表 2 土壤稳定入渗率 BP 神经网络缺省因子检验结果

缺省因子项	土壤稳定入渗率 (mm min <sup>-1</sup> )			敏感因子排序
	网络总误差 (mm min <sup>-1</sup> )	误差最大值 (mm min <sup>-1</sup> )	误差平均值 (mm min <sup>-1</sup> )	
全模型	0.005 1	0.183 7	2.23E-05	
降雨强度	0.012 9	0.264 3	2.81E-05	1
土壤前期含水量	0.005 9	0.216 4	-1.13E-05	3
坡度	0.004 7	0.181 5	2.25E-05	4
植被覆盖度	0.007 8	0.261 2	2.46E-05	2

## 4 土壤稳定入渗率的灰色关联分析

### 4.1 灰色关联分析

灰色关联分析是基于灰色系统理论的一种重要的多因素分析方法, 其基本思想是根据序列曲线几何形状的相似程度来判断其联系是否紧密, 曲线越接近, 相应序列之间的关联度就越大, 反之就越小。灰色关联分析弥补了传统数理统计方法在系统分析时带来的缺憾, 适用范围较为广泛<sup>[22]</sup>。设目标序列  $Y_0(p) = (Y_1(1), Y_1(2), \dots, Y_1(M))$ , 比较序列  $X_q(p) = (X_1(1), X_1(2), \dots, X_1(M))$ ,  $q = 1, 2, 3, \dots, N$ , 序列长度为  $M$ , 研究的问题集中在确定  $X_q(p)$  与  $Y_0(p)$  的密切程度, 也就是求  $X_q(p)$  与  $Y_0(p)$  序列的相关度。为确保结果的准确性, 首先对原始数据进行“中心化”无量纲处理<sup>[23]</sup>, 如式(6)所示:

$$X_q(p) = (x_q(p) - avx_q) / \sigma_q \quad (8)$$

式中,  $avx_q$  和  $\sigma_q$  分别为序列  $x_q(p)$  的平均值和均方差。

则  $x_q(p)$  和  $y_0(p)$  在点  $p$  的关联系数<sup>[22]</sup> 定义为:

$$\begin{aligned} \mu_q(p) &= \gamma(y_0(p), x_q(p)) \\ &= (\min_q \min_p |y_0(p) - x_q(p)| + \xi \min_q \min_p |y_0(p) - x_q(p)|) / (|y_0(p) - x_q(p)| + \xi \min_q \min_p |y_0(p) - x_q(p)|) \end{aligned} \quad (9)$$

式中,  $\mu_q(p)$  为第  $p$  时刻比较曲线  $x_q$  与参考曲线  $y_0$  的差值, 称为  $x_q$  与  $y_0$  在  $p$  时刻的关联系数;  $\xi$  为分辨系数, 一般有  $0 < \xi < 1$ , 本文认为被比较的两个序列同等重要, 故取  $\xi = 0.5$ 。

根据灰色关联空间定义可知, 序列  $x_q$  与  $y_0$  关联度为:

$$r_{oq} = (\sum r_{oq}(p)) / M \quad (10)$$

式中,  $p = 1, 2, 3, \dots, N$ 。

### 4.2 灰色关联分析结果

采用 BP 神经网络算法中的全部 51 组数据的输入样本作为比较序列, 输出样本作为参考序列进行灰色关联分析, 利用式(9)和式(10)计算出各因子与土壤稳定入渗率的关联度(表 3)。

表 3 土壤稳定入渗率与各因子灰色关联度

因子项	关联度 $r$	排序
降雨强度	0.899 2	1
植被覆盖度	0.757 2	2
土壤前期含水量	0.714 6	3
坡度	0.640 1	4

由表 3 可知, 各因子对土壤稳定入渗率的关联度排序为: 降雨强度 > 植被覆盖度 > 土壤前期含水量 > 坡度, 排序结果与 BP 模型缺省因子检验结果完全一致。通过对 51 组实测数据的偏相关分析发现, 降雨强度、植被覆盖度、土壤前期含水量和坡度与土壤稳定入渗率的偏相关系数分别为: 0.721 ( $\alpha = 0.000$ )、0.685 ( $\alpha = 0.000$ )、-0.644 ( $\alpha = 0.027$ ) 和 -0.462 ( $\alpha = 0.017$ )。由此证明了 BP 神经网络模型关于土壤稳定入渗率缺省因子敏感性分析的正确性, 同时也验证了灰色关联分析结果的准确性。

## 5 结 论

1) 通过改进后的 BP 算法建立了模拟土壤稳

定入渗率的 BP 神经网络模型,确定最优的网络拓扑结构为 4:10:1,此时,网络训练总误差为  $0.0024 \text{ mm min}^{-1}$ ,网络计算相对误差均值为 5.6%,是多元回归模型的 1/2,说明 BP 模型在定量化描述土壤入渗研究中具有更好的稳定性、准确性以及非线性映射能力。

2) 基于 BP 模型缺省因子检验得到各因子敏感性排序为:降雨强度 > 植被覆盖度 > 土壤前期含水量 > 坡度。灰色关联分析法对各因子的排序结果与 BP 模型缺省因子检验结果一致,进一步验证了 BP 模型的可靠性和准确性,同时也完成了灰色关联分析正确性的验证。

### 参 考 文 献

- [1] 黄昌勇. 土壤学. 北京: 中国农业出版社, 2000: 112—113
- [2] 雷志栋, 杨诗秀, 谢森传. 土壤水动力学. 北京: 清华大学出版社, 1988: 25—50, 77
- [3] Green W H, Ampt G A. Studies on soil physics: I. Flow of air and water through soils. *J Agr Sci*, 1911, 4(91): 1—24
- [4] Philip J R. The theory of infiltration. 1: The infiltration equation and its solution. *Soil Sci*, 1957, 83(5): 345—357
- [5] Smith R E. The infiltration envelope: Results from a theoretical infiltrometer. *Journal of Hydrology*, 1972, 17(1): 1—21
- [6] Smith R E, Parlange J Y. A parameter-efficient hydrologic infiltration model. *Water Res*, 1978, 14(3): 533—538
- [7] Kostikov A N. On the dynamics of the coefficient of water-percolation in soils and on the necessity for studying it from dynamic point of view for purposes of amelioration. *Trans Int Congr. International Society of Soil Science*, 1932, (A): 17—21
- [8] Horton R E. An approach toward a physical-interpretation of infiltration capacity. *Soil Sci Soc Am Proc*, 1940, 5: 399—417
- [9] 郭忠升, 邵明安. 半干旱区人工林地土壤入渗过程分析. *土壤学报*, 2009, 46(5): 953—958
- [10] 周蓓蓓, 邵明安. 不同碎石含量及直径对土壤水分入渗过程的影响. *土壤学报*, 2007, 44(5): 801—807
- [11] 李毅, 王全九, 邵明安, 等. Green-Ampt 入渗模型及其应用. *西北农林科技大学学报: 自然科学版*, 2007, 35(2): 225—230
- [12] 张振华, 谢恒星, 刘继龙, 等. 基于图形特征的 Green-Ampt 入渗模型关键参数  $S_f$  和  $K_s$  的简化求解. *土壤学报*, 2006, 43(2): 203—208
- [13] 赵景波, 张允, 陈宝群, 等. 陕西洛川中更新统下部黄土入渗规律研究. *土壤学报*, 2009, 46(6): 965—972
- [14] 陈守余. 神经网络模拟实现与应用. 武汉: 中国地质大学出版社, 2000: 1—6
- [15] 韩力群. 神经网络教程. 北京: 北京邮电大学出版社, 2006: 49—50
- [16] 韩力群. 神经网络理论、设计及应用——人工神经细胞、神经网络和人工神经系统. 北京: 化学工业出版社, 2002: 50
- [17] 冯绍元, 霍再林, 康绍忠, 等. 干旱内陆区自然——人工条件下地下水水位动态的 ANN 模型. *水利学报*, 2007, 38(7): 873—878
- [18] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用. 合肥: 中国科学技术大学出版社, 1998: 65—66
- [19] 杨建刚. 神经网络实用教程. 杭州: 浙江大学出版社, 2001: 47—48
- [20] 刘延峰, 靳孟贵, 曹英兰. BP 网络在焉耆盆地土壤积盐影响因素分析中的应用. *中国农村水利水电*, 2004(2): 20—24
- [21] 刘国东, 丁晶. BP 网络用于水文预测的几个问题探讨. *水利学报*, 1999(1): 65—70
- [22] 刘思锋, 郭天榜, 党耀国, 等. 灰色系统理论及其应用. 北京: 科学出版社, 1999: 43—49
- [23] 李炳军, 朱春阳, 周杰. 原始数据无量纲化处理对灰色关联序的影响. *河南农业大学学报*, 2002, 36(2): 199—202