

江西退化红壤区湿地松树干液流 BP神经网络模型研究*

涂洁

(南昌工程学院生态与环境科学研究所, 江西 南昌 330099)

摘要: 以我国南方亚热带退化红壤区重要植被恢复树种——湿地松 *Pinus elliotii* 为研究对象, 采用相关分析法选定空气温度、空气相对湿度、平均净辐射、水汽压亏缺为输入变量, 以液流速率为输出变量, 构建拓扑结构为 4-10-1 的湿地松树干液流 BP 神经网络模型。采用贝叶斯正则化法和 Levenberg-Marquardt 法对 1 900 组实测数据进网络训练, 用剩余的 1 900 组实测样本进行检验。结果表明: 在两种算法下, 训练样本和检验样本模型输出值与实测值之间线性回归的拟合程度均较高, 回归方程的相关系数达 0.98。训练样本的拟合精度分别为 88.12% 和 88.11%, 检验样本的仿真精度分别为 88.11% 和 87.98%。该模型对树干液流速率具有较高的模拟精度和泛化能力, 能够很好地反映气象因子与液流速率之间的非线性函数关系, 从而为同类条件下湿地松人工林的可持续经营和林地水资源的科学管理提供一种新方法。

关键词: 湿地松 *Pinus elliotii*; 树干液流; BP 神经网络; 贝叶斯正则化算法; Levenberg-Marquardt 算法

中图分类号: S718.43 **文献标志码:** A **文章编号:** 0529-6579 (2016) 01-0024-06

BP neural network models of sap flow velocity for *Pinus elliotii* in degraded red soil area of Jiangxi Province

TU Jie

(Research Institute of Ecology & Environmental Sciences Nanchang Institute
of Technology, Nanchang 330099, China)

Abstract: *Pinus elliotii* is commonly considered to be one of the most important pioneer tree species for forest restoration and reconstruction in subtropical degraded red soil area of China, due to its high productivity and leanness-resistance. However, they also consume certain amount of water during the course of growing and maintaining ecological balance. Therefore, quantitative research on tree water consumption by transpiration has become a hotspot in the field of tree physiological ecology in recent years. In order to provide an effective way for conducting sustainable management of *Pinus elliotii* plantation and associated water resource of similar condition, in this study, we chose the *Pinus elliotii* plantation in degraded red soil area of Jiangxi province as the research object, and the log-sigmoid type function (i. e. tansig) of MATLAB software was selected as the action function of neurons. Based on the correlation analysis between sap flow velocity and meteorological factors, air temperature, relative air humidity, average net radiation and vapor pressure deficit were chosen as the input variables and sap flow velocity as the output variable. Then the optimum 3-layer BP artificial network model of individual sap flow velocity was established with the topological structure of 4-10-1. Nineteen hundreds groups of individual tree data were

* 收稿日期: 2014-07-22

基金项目: 国家自然科学基金资助项目 (31260172); 江西省教育厅科学技术研究资助项目 (GJJ14753)

作者简介: 涂洁 (1980 年生), 女; 研究方向: 区域生态生产力; E-mail: tujie8058@163.com

used to train the very neutral network both with Bayesian regularization algorithm and Levenberg-Marquardt algorithm, while the remaining nineteen hundreds groups were used to test the model. Good fitting results were obtained for linear regression between predictive and measured values under two algorithm model, with the R higher as 0.98. The results showed that fitting accuracy of training samples was 88.12% and 88.11%, respectively, and the simulating accuracy of testing samples was 88.11% and 87.98%, respectively. This model can well reflect the non-linear relationship between sap flow velocity and meteorological factors owing to its higher accuracy and fine generalization than linear regression model, which indicated the availability of BP neural network for the analysis, simulation and prediction of sap flow velocity.

Key words: *Pinus elliottii*; sap flow; BP neural network; Bayesian regularization algorithm; Levenberg-Marquardt algorithm

液流是植物的一个重要生理生态参数，可反映植物体内的水分传输状况^[1]、植物对水分的利用特征^[2]及其对环境的响应^[3-4]，也是验证、修正蒸腾（蒸散）模型的重要参数^[5]。液流与气象因子间的关系也因此成为当前植物生理生态的研究热点之一，并建立了描述两者之间关系的多元线性回归模型^[6-8]。但这些模型通常假定生物与环境的关系都是平滑、连续的线性多项式关系，不能全面反映其非线性关系。BP 神经网络是采用 BP 算法训练权值的多层前馈网络，可以很好地解决传统概率函数模型适用范围窄、模型实用性差等问题，非常适合解决林业问题。Foody G M^[9]、Susan^[10]通过对 BP 神经网络与其它方法模拟效果的比较发现，BP 神经网络更具优势。目前 BP 神经网络在林业领域的应用研究大多围绕林木生物量^[11]、森林碳蓄积^[12]、林分直径分布^[13]以及景观生态规划^[14]等方面展开，而利用人工神经网络构建树干液流模型仍处于探索阶段。朱建刚等^[15]成功利用 BP 神经网络构建了油松、侧柏的液流预测模型，得到的网络模型预测精度较高，能够用于实际预测。

本文以我国南方亚热带退化红壤区重要植被恢复树种——湿地松为研究对象，以热扩散探针法对树干液流的测定结果为目标向量，以其主要影响因子——空气温度（Ta）、空气相对湿度（RH）、平均净辐射（ANR）、水汽压亏缺（VPD）为输入变量，在 MATLAB 7.0 软件平台上利用嵌套的贝叶斯正则化法和 L-M 法进行网络训练，构建湿地松树干液流 BP 神经网络模型，探索树干液流速率与其主要气象要素间的非线性函数关系映射，并对网络的预测精度和泛化能力进行检验，为退化红壤区湿地松人工林的科学经营和林地水资源的有效管理提供一种新的技术方法。

1 研究方法

1.1 试验区概况

试验地点位于江西省泰和县中国科学院千烟洲试验区（115°04'13"E, 26°44'48"N）。区内年平均气温 17.9℃，≥ 0℃ 活动积温 6 253℃，≥10℃ 活动积温 6 015℃，多年平均降水量 1 489 mm，年日照时数 1 406 h，年日照百分率 43%，太阳年总辐射量 4 349 MJ/m²，无霜期 323 d，具有典型的亚热带季风气候特征。样地设在 28 年生湿地松人工纯林内，2012 年 4 月调查得到乔木层平均胸径 22.3 cm，平均树高 16.3 m，平均密度约 1 760 株·hm⁻²。由于林分基本郁闭，下木和草本数量较少，灌木层主以白栎 *Quercus fabri*、牡荆 *Vitex negundo* 等为主。草本层数量更少，其代表性类型为刺芒野古草 *Aruudinella steosa*。

1.2 样本数据采集

根据 TDP 探头的长度和被测木具有代表性的原则，选取生长良好、树干通直、无被挤压的湿地松 3 株[胸径(22.3 ± 0.8) cm]进行树干液流测定。在被测木树干 1.3 m 处安装 TDP 探针（型号 TDP-30, Dynamax 公司, 美国），另一端与数据采集器（DT-80, Data Taker 公司, 澳大利亚）连接。采用自动气象站记录空气温度（Ta）、空气相对湿度（RH）、平均净辐射（ANR）等环境因子（数据采集间隔 30 min）。采用水汽压亏缺（VPD）综合表达空气温度与空气相对湿度的协同效应，计算方法见文[16]。

树干液流速率 v (cm · s⁻¹) 由 Granier 经验公式计算得到。

$$v = 0.0119 \times \left(\frac{\Delta T_{\max} - \Delta T}{\Delta T} \right)^{1.231}$$

式中， ΔT 为两探针间的温差， $\Delta T_{\max}$ 为连续 7 ~ 10 d 所测液流数据中的最大值^[17]。

2 模型构建与检验

在 BP 神经网络建模过程中，容易出现样本冗余信息、收敛于局部极小点、网络结构不合理、网络泛化能力差等问题。本文在建模过程中针对这些关键问题进行了改进。

2.1 样本数据预处理

选取 2012 年 4 - 12 月湿地松液流速率与气象因子共 5 120 组数据作为模型的初始样本数据集。由于网络训练前的数据处理对网络有至关重要的影响，可影响网络的学习速度、网络结构的复杂性和网络的精度。主要包括：

1) 在数据采集过程中，设备由于外界干扰或人为操作失误会导致数据缺失或异常，因此需要将含有缺失或异常数据的行从初始样本集中删除。文中采用李海涛等^[18]在相关文献中提到的无量纲判别参数 P 判断试验监测数据中出现的异常值。

2) 利用 SPSS 软件中 Pearson 双尾相关分析法对典型月份液流速率与相关气象因子进行偏相关性

分析，结果发现液流速率与平均净辐射、空气温度、水汽压亏缺呈显著正相关，与空气相对湿度呈显著负相关（见表 1）。

3) 由于采集数据中各因素、指标的量纲或数量级不同，为了更好地反映各因素间的相互关系，充分发挥 BP 神经网络的预测功能，提高其输出精度，模型训练之前需对具体对象真实的数据进行标准化处理。由于 BP 网络通常采用的神经元作用函数是 S 形函数，以双曲正切函数 tansig 函数为例，tansig 函数在 $[-1, 1]$ 这一区间具有最大斜率，因此需要将表 2 中的数据都归一化到 $[-1, 1]$ 范围内，文中通过 Matlab 工具箱中提供的 premnmx 和 postmnmx 函数实现。计算公式为：

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

式中：X、X’ 分别为变量归一化前、后的数值； $X_{\max}$ 为某一变量中的最大值； $X_{\min}$ 为某一变量中的最小值。

表 1 湿地松液流速率与气象因子的偏相关分析¹⁾
Table 1 Partial correlation analysis between sap flow velocity of *Pinus elliottii* and meteorological factors

控制变量	分析变量	偏相关系数		
		4 月	7 月	10 月
Ta、RH、VPD	ANR	0.806**	0.775**	0.473**
RH、VPD、ANR	Ta	0.642**	0.384**	0.419**
Ta、ANR、VPD	RH	-0.455**	-0.748**	-0.288**
Ta、ANR、RH	VPD	0.354**	0.698**	0.419**

1) ** 为 0.01 水平上显著相关

表 2 湿地松液流预测初始样本数据集
Table 2 Initial sample data set of *pinus elliottii* for sap flow forecasting

时 间	液流	冠层温度/℃	平均净辐射/(W·m ⁻²)	层顶平均相对湿度/%	水汽压亏/kPa
2012-4-20 10:00	0.000 558	18.770	22.364	102.42	-0.030 607 5
2012-4-20 11:00	0.000 593	22.825	223.23	102.29	-0.034 679 1
...	...	...	...	...	...
2012-12-31 12:00	0.000 296	13.511	75.033	55.345	0.414 944 4

2.2 网络结构参数的确定

网络结构参数包括网络的层数、每层的节点数、每层神经元的作用函数、训练方法和学习速率等。文中采用三层 BP 网络，包括输入层、输出层和隐含层。由于模型是将 4 个气象因子作为输入变量，液流速率作为输出变量，输入层选择四个节点，输出层选择一个节点。确定隐含层节点数的基

本原则是：在满足精度要求的前提下取尽可能少的节点数，建立紧凑的网络结构。试凑法通过先设置较少的节点数对网络进行训练，并测试网络的逼近误差，然后逐渐增加节点数，直到测试误差不再有明显的减小为止。“试凑法”训练发现隐含层节点数为 10 左右时网络的逼近效果最好。因此，树干液流 BP 神经网络模型的拓扑结构确定为 4 - 10 -

1. 前期采用 SPSS 软件对数据进行预处理时发现，当采用线性模型结构对数据进行拟合时无法得到理想的收敛结果，因此本文每层神经元的作用函数均采用 tansig 函数，以期拟合出较为复杂的非线性映射关系。网络的学习速率设置为 0.1，最小误差 0.000 1，迭代次数 1 000。

为提高网络模型的泛化能力，文中同时采用 MATLAB 工具箱中提供的贝叶斯正则化方法和 L-M 方法（Levenberg—Marquardt）。贝叶斯正则化方法和 L-M 方法都是神经网络训练中常用的方法，其中 L-M 法由于避免了直接计算赫赛矩阵，减少了训练中的计算量，对于中等规模的 BP 神经网络有最快的收敛速度。相对 L-M 训练方法，贝叶斯正则化方法对噪音更不敏感，具有更好的泛化能力，但训练速度相对较慢。

2.3 模型的检验与分析

将经预处理后的 3 800 组有效数据随机分为两组，其中 1 900 组作为训练样本，余下的 1 900 组作为检验样本。模型预测精度的检验采用测试样本的模拟输出与期望（目标）输出之间的线性回归拟合程度实现。模型泛化能力的检验采用训练样本回归检验与测试样本回归检验进行比较得出，当测试样本与训练样本的模型输出与实际检测结果线性回归的拟合程度相当或稍大时，表明模型的泛化能力较好。

2.3.1 模型预测精度的检验 将实测值作为被解释变量，预测值作为解释变量，建立一元线性回归方程 $y = ax + b$ 。如果模型预测值与实测值线性关系显著、回归方程合理，则 a 越接近 1， b 值越接近 0，说明预测值与实测值之间吻合度越高。图 1 为液流速率实测值与预测值之间的线性回归分布情况。图 1 不难看出，液流速率 BP 网络模型的输出结果与实测结果在上述 4 种情况下都较好的符合了 $y = x$ 的分布。

对 BP 网络模型总体训练结果进行分析，具体情况见表 3。线性回归方程的参数估计按照公式（1）计算得到，相关系数由公式（2）计算得到。从表 3 中可以看出，当隐含层节点数选择为 10 左右时，无论是训练样本还是测试样本，采用贝叶斯正则化方法和 L-M 法都能得到较好的线性回归结果，回归方程的相关系数均在 0.93 以上。运用两种算法对训练样本进行网络训练，拟合精度分别为 88.12% 和 88.11%，将检验样本代入液流 BP 神经网络模型进行检验，仿真精度分别为 88.11% 和 87.98%。

$$\hat{b} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}, \quad \hat{a} = \bar{y} - \hat{b} \bar{x} \tag{1}$$

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{2}$$

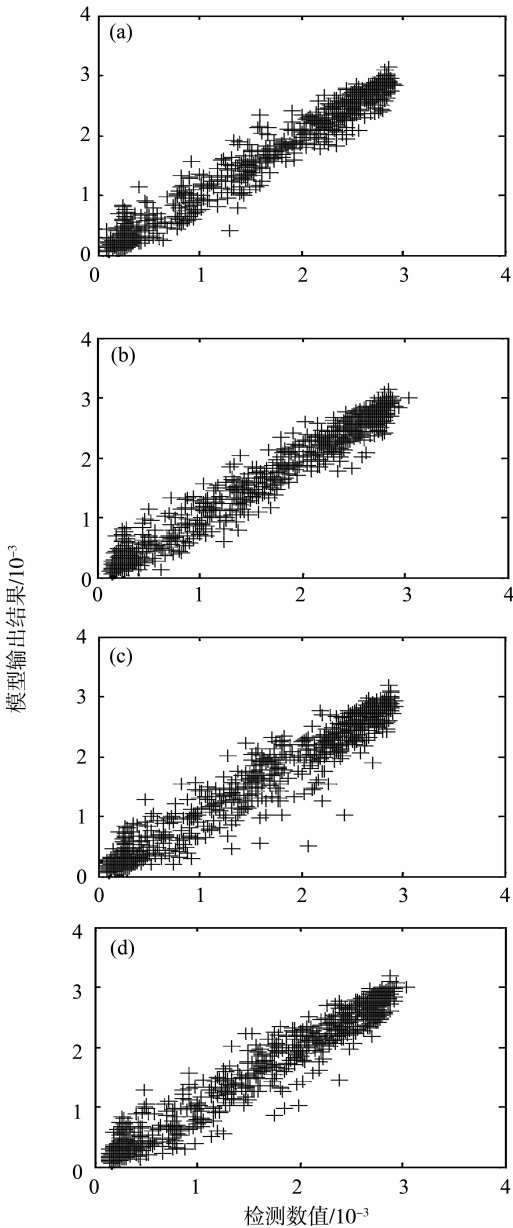


图 1 液流速率实测值与预测值线性回归情况
Fig.1 Regression distribution of sap flow measured and predictive value

- (a) 为训练样本贝叶斯正则化训练得到的结果；
- (b) 为训练样本 L-M 方法训练得到的结果；
- (c) 为检验样本贝叶斯正则化训练得到的结果；
- (d) 为检验样本 L-M 方法训练得到的结果

表 3 BP 网络模型训练结果统计情况
Table 3 Statistics of training results of BP neural network model

训练模式	各层最佳节点数	训练样本线性回归分析	测试样本线性回归分析
贝叶斯正则化法	4—10—1	$y = 0.983\ 17x + 0.000\ 117$ $R = 0.976\ 77$ Sig. <0.001	$y = 0.991\ 42x + 0.000\ 001\ 173\ 2$ $R = 0.981\ 44$ Sig. <0.001
L - M 法	4—10—1	$y = 0.999\ 04x + 0.000\ 001\ 949\ 9$ $R = 0.986\ 42$ Sig. <0.001	$y = 1.000\ 4x + 0.000\ 000\ 320\ 3$ $R = 0.985\ 96$ Sig. <0.001

2.3.2 模型泛化能力的检验 图 2 为样本数据集中随机 24 h 液流速率模型输出结果与实测结果，图 2 表明，运用两种算法进行网络训练得出的拟合值、仿真值均能很好地逼近实测值，由于训练样本对网络模型参数的直接影响，利用训练样本直接测试得到的相关系数和拟合精度均稍高于测试样本的结果。因此，该 BP 网络模型具有较好的泛化能力。

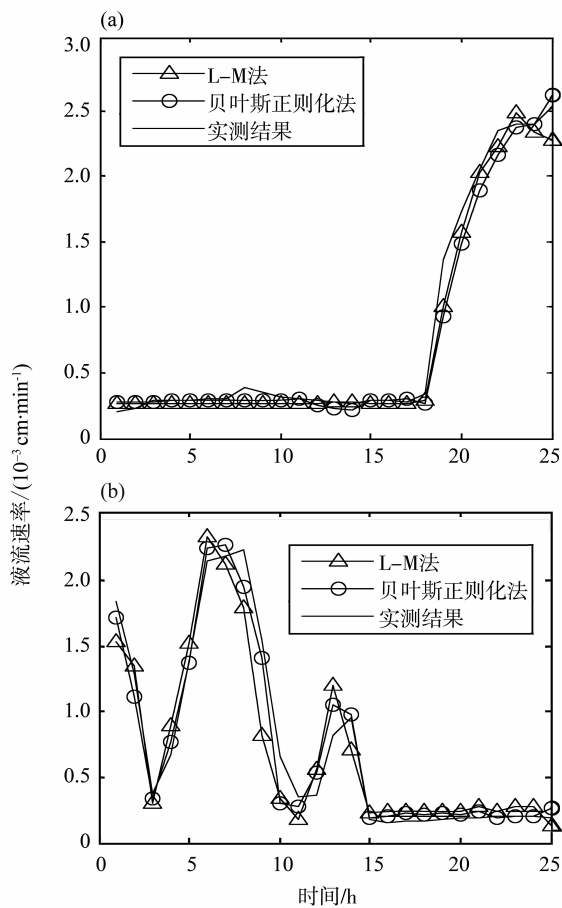


图 2 液流速率实测值与预测值比较验证
Fig. 2 Comparison test of sap flow velocity measured and predictive value

(a) 为训练样本模型输出与实测结果拟合情况；
(b) 为测试样本模型输出与实测结果拟合情况

3 讨论与结论

1) 利用 SPSS 相关分析法得出平均净辐射、空气温度、水汽压亏缺、空气相对湿度为输入项，液流速率为目标输出项，采用贝叶斯正则化和 L-M 两种算法进行网络训练，建立拓扑结构为 4 - 10 - 1 的树干液流 BP 神经网络模型。从模型的检验结果看出，训练样本与测试样本的模型输出与实测结果的线性回归拟合程度相当，回归方程的相关系数达 0.98，拟合精度均达 88%。因此，两种算法下建立的 BP 神经网络模型均表现出较高的预测精度和网络泛化能力，实现了精度与实用的统一，可以作为对传统树干液流速率建模方法的一种有益补充。

2) 由于本研究是在特定地域条件下进行的，选取样本的试验地范围较小，因此所建立的 4 因子树干液流 BP 神经网络模型具有一定的局限型，仅在建模区域有效，没有普适性。能否可以应用类似方法进行更大尺度、更复杂林分的耗水规律模拟，需要进一步的思考研究。因此，对于不同地区、不同树种以及林分因子和经营要素更多因素的考虑是今后探讨的方向。

3) 虽然神经网络非线性映射能力很好，但其固有的缺点需要不断改进，目前已有一些新的算法被用来改进网络的训练过程，如泛化改进算法^[19]、WASD 算法^[20]、遗传算法^[21]。此外，朱建刚等^[15]研究认为虽然应用 BP 神经网络建立的植物液流预测模型具有较好的预测性能，但神经网络为特有的“黑箱”式构建过程无法直观定量不同因素与目标值之间的依赖关系^[22]，不能解释液流传输的机制，因此需要进行更为广泛深入的研究探讨。

致谢：感谢中国科学院地理科学与资源研究所生态系统网络观测与模拟重点实验室千烟洲生态站为本论文基础数据提供帮助。

参考文献:

- [1] COHEN M, GOLDHAMER D, FERERES E, et al. Assessment of peach tree responses to irrigation water deficit by continuous monitoring of trunk diameter changes[J]. The Journal of Horticultural Science and Biotechnology, 2001, 76: 55 – 60.
- [2] WULLSCHLEGER S D, MEINZER F C, VERTESSY R A. A review of whole-plant tree water use studies in trees [J]. Tree Physiology, 1998, 18: 499 – 512.
- [3] KELLOMÄKI S, WANG Kaiyun. Sap flow in Scots pines growing under conditions of year-round carbon dioxide enrichment and temperature elevation [J]. Plant, Cell and Environment, 1998, 21: 969 – 981.
- [4] WULLSCHLEGER S D, NORBY R J. Sap velocity and canopy transpiration in a sweet gum stand exposed to free-air CO₂ enrichment (FACE) [J]. New Phytologist, 2001, 150: 489 – 498.
- [5] GONG Daozhi, KANG Shaozhong, ZHANG Lu, et al. A two-dimensional model of root water uptake for single apple trees and its verification with sap flow and soil water content measurements [J]. Agricultural Water Management, 2006, 83: 119 – 129.
- [6] 赵仲辉, 康文星, 田大伦, 等. 湖南会同杉木液流变化及其与环境因子的关系[J]. 林业科学, 2009, 45(7): 127 – 132.
- [7] 王文杰, 孙伟, 邱岭, 等. 不同时间尺度下兴安落叶松树干液流密度与环境因子的关系[J]. 林业科学, 2012, 48(1): 77 – 84.
- [8] 池波, 蔡体久, 满秀玲, 等. 大兴安岭北部兴安落叶松树干液流规律及影响因子分析[J]. 北京林业大学学报, 2013, 35(4): 21 – 26.
- [9] FOODY G M, BOYD D S, CUTLER M L. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions [J]. Remote Sensing of Environment, 2003, 85(4): 463 – 474.
- [10] SUSAN L, KING K P, BENNETT S L. Modeling non-catastrophic individual tree mortality using logistic regression neural networks and support vector methods [J]. Computers and Electronics in Agriculture, 2000, 27(1–3): 401 – 406.
- [11] 王轶夫, 孙玉军, 郭孝玉. 基于 BP 神经网络的马尾松立木生物量模型研究[J]. 北京林业大学学报, 2013, 35(2): 17 – 21.
- [12] 许炜敏, 陈友飞, 陈明华, 等. 基于 BP 神经网络的杉木林蓄积量估测研究[J]. 福建林学院学报, 2012, 32(4): 310 – 315.
- [13] 黄家荣, 孟宪宇, 关毓秀. 马尾松人工林直径分布神经网络模型研究[J]. 北京林业大学学报, 2006, 28(1): 28 – 31.
- [14] 郭烁, 夏北成, 江学顶. 基于 GIS 与人工神经网络的广州森林景观生态规划[J]. 中山大学学报(自然科学版), 2005, 44(5): 121 – 123.
- [15] 朱建刚, 余新晓, 陈丽华. 预测植物瞬态液流的 BP 神经网络模型[J]. 林业科学, 2010, 46(1): 152 – 157.
- [16] CAMPBELL G S, NORMAN J M. A introduction to environmental biophysics [M]. 2nd ed. US: Springer-verlag Press, 1998: 40 – 45.
- [17] LU Ping, URBAN L, ZHAO Ping. Granier's Thermal dissipation probe (TDP) method for measuring sap flow in trees: Theory and practice [J]. Acta Botanica Sinica, 2004, 46(6): 631 – 646.
- [18] 王轶夫, 孙玉军, 郭孝玉. 基于 BP 神经网络的马尾松立木生物量模型研究[J]. 北京林业大学学报, 2013, 35(2): 17 – 21.
- [19] 琚存勇, 蔡体久. 用泛化改进的 BP 神经网络估测森林蓄积量[J]. 林业科学, 2006, 42(12): 59 – 62.
- [20] 张雨浓, 王茹, 劳稳超, 等. 符号函数激励的 WASD 神经网络与 XOR 应用[J]. 中山大学学报(自然科学版), 2014, 53(1): 1 – 7.
- [21] 桂凌, 张征, 闫国振, 等. 基于遗传神经网络的鄂尔多斯沙棘水土保持功能评价[J]. 干旱区资源与环境, 2012, 26(7): 136 – 140.
- [22] 杨娟, 王昌全, 李冰, 等. 基于 BP 神经网络的城市边缘带土壤重金属污染预测—以成都平原土壤 Cd 为例[J]. 土壤学报, 2007, 44(3): 430 – 436.