

# 基于 BP 神经网络预测林内 PM<sub>2.5</sub> 浓度

陈博<sup>1</sup>, 李迎春<sup>2</sup>, 夏振平<sup>1</sup> (1. 北京农业职业学院, 北京 102442; 2. 黄堡苗圃, 北京 102604)

**摘要** [目的]利用 BP 神经网络预测林内 PM<sub>2.5</sub> 浓度。[方法]利用神经网络理论,采用 2013 年 7 月—2014 年 5 月野外实时监测数据,建立了以气象参数、污染源强变量和林分结构特征为输入因子,林内 PM<sub>2.5</sub> 小时平均浓度为输出因子的预测模型,并对其预测精度进行了评价。[结果]BP 神经网络模型能够很好地捕捉污染物浓度与气象因素和林分结构间的非线性影响规律,预测结果的平均相对误差为  $1.71 \times 10^{-3}$ , 均方根误差为 6.77, 拟合优度达 0.98, 模型具有很高的预测精度。而传统的多元线性回归 (MLR) 模型预测结果的平均相对误差、均方根误差和拟合优度分别为 0.27、22.92 和 0.93。[结论]研究成果印证了应用 BP 神经网络模型预测林内 PM<sub>2.5</sub> 浓度的可行性和准确性。

**关键词** PM<sub>2.5</sub>; BP 神经网络; 多元线性回归; 林分结构

**中图分类号** S771.8 **文献标识码** A

**文章编号** 0517-6611(2019)01-0107-04

**doi**: 10.3969/j.issn.0517-6611.2019.01.033



开放科学 (资源服务) 标识码 (OSID):

## Prediction of PM<sub>2.5</sub> Concentration in Forest Based on BP Artificial Neural Network

CHEN Bo<sup>1</sup>, LI Ying-chun<sup>2</sup>, XIA Zhen-ping<sup>3</sup> (1. Beijing Vocational College of Agriculture, Beijing 102442; 2. Huangfa Nursery, Beijing 102604)

**Abstract** [Objective] BP neural network was used to predict PM<sub>2.5</sub> concentration in forest. [Method] An artificial neural network model was established with meteorological data, atmospheric PM<sub>2.5</sub> concentration outside the forest and forest structure as the input factors, and PM<sub>2.5</sub> hourly average concentration inside the forest as the output factors. Its prediction accuracy was also evaluated in this paper. [Result] BP artificial neural network model can be trained to model the highly non-linear relationships between meteorological parameters, forest structure and PM<sub>2.5</sub> concentration. Mean relative error ( $E_{MR}$ ), root mean square error ( $E_{RMS}$ ) and goodness of fit ( $R^2$ ) between BP simulated PM<sub>2.5</sub> concentrations in the forest and observed ones were  $1.71 \times 10^{-3}$ , 6.77 and 0.98, respectively. The mean relative error, root mean square error and goodness of fit of traditional MLR model were 0.27, 22.92 and 0.93, respectively. [Conclusion] It can be concluded that BP artificial neural network model is a promising approach in predicting PM<sub>2.5</sub> concentration inside a forest.

**Key words** PM<sub>2.5</sub>; BP artificial neural network; Multiple linear regression; Forest structure

PM<sub>2.5</sub> 指大气中直径小于或等于 2.5 μm 的颗粒物,常被称为可入肺颗粒物或细颗粒物<sup>[1]</sup>。虽然 PM<sub>2.5</sub> 只是大气成分中含量很少的组分,但其在空气中的重力沉降率几乎为零,长期在空气中悬浮,附着很多有机污染物和重金属元素<sup>[2]</sup>,对人类健康、环境和气候有重要影响。植物作为环境绿化的主体,对空气有一定的净化作用,森林植被可以通过直接和间接的方式对大气中颗粒物的浓度产生影响<sup>[3-6]</sup>。然而森林植被对 PM<sub>2.5</sub> 的作用研究尚处于起步阶段,林内 PM<sub>2.5</sub> 浓度变化规律及林内外浓度对比情况尚不清楚。已有文献报道林内 PM<sub>2.5</sub> 的质量浓度受污染散发源、气象条件和林分结构等多种因素影响<sup>[7-10]</sup>,既存在林内 PM<sub>2.5</sub> 浓度低于林外,也存在林内 PM<sub>2.5</sub> 浓度高于林外的情况。鉴于此,准确预测林内 PM<sub>2.5</sub> 浓度对于研究森林调控 PM<sub>2.5</sub> 的作用效果及引导公众健康出游、亲近森林具有重要意义。

研究表明,PM<sub>2.5</sub> 浓度的变化与气象条件之间呈很强的非线性关系<sup>[11]</sup>,传统的多元线性回归模型预测 PM<sub>2.5</sub> 质量浓度存在很大局限性。然而神经网络能够建立非常复杂的非线性模型,很好地反映 PM<sub>2.5</sub> 质量浓度与参数之间的关系<sup>[12-13]</sup>,此种方法已经成功地用于多种污染物的预测研究<sup>[14-15]</sup>。笔者以北京市大兴区黄村镇景观生态林为研究对象,应用 BP 神经网络模型预测林内 PM<sub>2.5</sub> 小时平均质

量浓度,并与多元线性回归预测结果做比较,说明基于神经网络模型预测林内 PM<sub>2.5</sub> 质量浓度的可行性和准确性。

## 1 实验方法原理

BP 神经网络属于多层前馈神经网络,据统计,有 80%~90% 的神经网络模型采用了 BP 网络或它的变化形式<sup>[16]</sup>。BP 神经网络的构成包括 1 个输入层、1 个或多个隐藏层以及 1 个输出层,主要特点为信号前向传递,而误差反向传播。BP 算法的原理是在前向的传递中,信号经由输入层、隐含层和输出层的逐层处理及传递,如果预测输出和给定输出之间的误差未达到精度要求,则转为反向传播并修正连接权值,进而促使预测输出逐渐逼近期望输出。3 层 BP 神经网络的结构如图 1 所示。一般采用 S 型的转换函数作为传递输入层与隐含层之间关系的函数,采用纯线性的转换函数作为传递隐含层与输出层之间关系的函数<sup>[17]</sup>。BP 神经网络输入层和隐含层间的 S 型传递函数表达公式如下:

$$F(x) = 1 / (1 + e^{-x}) \quad (1)$$

式中,  $x$  表示输入该神经元,将神经元  $(-\infty, +\infty)$  的输入范围映射到  $(0, +1)$  较小的范围内,便于 BP 算法训练神经网络。

## 2 实验设计

**2.1 数据获取** 该研究中采用的 283 组数据由野外实时监测得到。2013 年 7 月—2014 年 5 月,每个季度随机抽取无降水天气 9 d,每天连续监测 9 h。在北京市大兴区黄村镇大洼村景观生态林林内监测点和林外 2 个监测点同步监测 1.5 m 高处空气中的 PM<sub>2.5</sub> 浓度及气象要素 (温度、相对湿度、风速、风向),为了保证监测点具有相近的污染源,且避免监测

**基金项目** 国家林业公益性行业科研项目 (201304301)。

**作者简介** 陈博 (1983—),女,吉林大安人,讲师,博士,从事园林生态研究。

**收稿日期** 2018-08-31

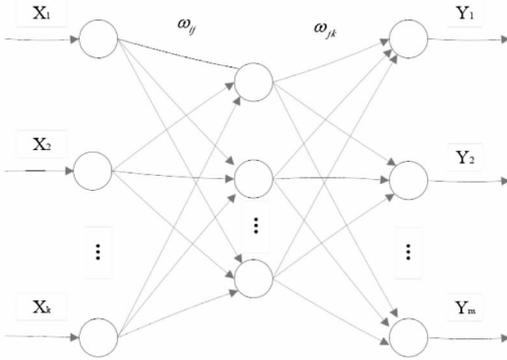


图1 3层BP神经网络的结构

Fig. 1 Structure of three-layer BP neural network

点出现林缘效应,林内监测点设在景观生态林中心位置,距林缘 50 m,林外监测点设在林外距林缘 50 m 的位置且四周空旷无高大乔木和建筑遮挡。林内样地主要乔木为旱柳 (*Salix matsudana*),平均树高 6 m,平均胸径 12 cm,林缘伴生榆叶梅 (*Prunus triloba*)和碧桃 (*Prunus persica*)等少量的花灌木。

采用英国 Turnkey 仪器制造公司生产、符合粉尘监测国家标准的 Dustmate 粉尘检测仪(分辨率  $0.1 \mu\text{g}/\text{m}^3$ ,测量范围  $0\sim 6\,000 \mu\text{g}/\text{m}^3$ ,粒径范围  $0.5\sim 15.0 \mu\text{m}$ )测定  $\text{PM}_{2.5}$  浓度,使用台湾衡欣 az8918 风速气温湿度三合一测试仪及 TN-F 三杯式风速表测定气温、相对湿度、风速和风向等。

**2.2 数据处理** 由于大气污染物的浓度变化受气象条件的影响,因此要实现对林内  $\text{PM}_{2.5}$  质量浓度的预测,必须同时考虑林外  $\text{PM}_{2.5}$  背景值及林内气象条件。

(1)背景浓度:将林外监测点的  $\text{PM}_{2.5}$  浓度设为预测模型输入参数中的背景浓度,即用 PM 表征污染源强对林内  $\text{PM}_{2.5}$  浓度的影响。

(2)气象参数:选取温度( $T$ )、相对湿度( $H$ )和风速( $V_w$ )数据为气象参数直接进入分析。风向( $\theta$ )经由正弦和余弦转换成 2 个变量  $\sin\theta$  和  $\cos\theta$ <sup>[17]</sup>,即把风向转换成为东西方向和南北方向 2 个变量。

(3)季节参数:考虑一年中不同季节植物生长快慢不同对林内  $\text{PM}_{2.5}$  浓度变化产生不同影响,引进季节变量  $A_{\text{season}}$  作为双变量使用。根据旱柳生长习性,春季和夏季生长旺盛,令  $A_{\text{season}}=0$ ;秋季和冬季生长缓慢,令  $A_{\text{season}}=1$ 。

建立的预测模型中全部的输入变量如表 1 所示。

表 1 预测模型输入变量

Table 1 Input variables of prediction model

输入变量 Input variable	变量名称 Variable name	变量符号 Variable symbol	单位 Unit
背景浓度 Background concentration	$\text{PM}_{2.5}$ 背景浓度	$\text{PM}_{2.5}^0$	$\mu\text{g}/\text{m}^3$
气象参数 Meteorological parameter	温度	$T$	$^{\circ}\text{C}$
	相对湿度	$H$	%
	风速	$V_w$	$\text{m}/\text{s}$
	风向正弦	$\sin\theta$	
	风向余弦	$\cos\theta$	
季节变量 Seasonal variable	季节	$A_{\text{season}}$	

**2.3 BP 人工神经网络模型** 基于 BP 人工神经网络方法仿真林内  $\text{PM}_{2.5}$  浓度的过程主要由 Matlab R2013a 中神经网络工具箱所提供的函数开展编程并建立起相应的模型而实现。具体网络实现步骤如下:Step1,网络初始化。将输入层到隐含层以及隐含层到输出层所有的权值设置成任意小的随机数,并设定初始阈值。Step2,读入输入变量与输出变量(林内  $\text{PM}_{2.5}$  质量浓度)。全部样本数据被随机分成训练组和测试组:训练组样本占 70%用于训练网络,根据显示的误差调试网络,其中训练法则为 Levenberg-Marquardt 运算法则;测试组样本占 30%用于独立测试训练网络的性能,结果对训练过程无影响。Step3,选择训练函数,在设置隐含层神经元的数目后训练网络。当网络不能收敛时,则需要更换训练函数;当结果误差比较大时,则需要将隐含层神经元的数目进行调整,之后再次训练网络,如此反复训练,最终获得最佳训练结果。Step4,确定 BP 神经网络的最终结构:单隐层,各层神经元数目为 7-6-1,训练函数确定为 trainlm,并使用训练好的网络对林内  $\text{PM}_{2.5}$  小时平均浓度进行仿真。

**2.4 多元线性回归模型** 将多元线性回归(multiple linear regression, MLR)模型与人工神经网络模型相比较,其中 MLR 模型是通过 Matlab R2013a 所建立。MLR 模型的数学表达式如下:

$$Y = b_0 + \sum_{i=1}^n b_i X_i + \varepsilon_i \quad (2)$$

式中, $X_i$  表示输入变量  $i$  的值; $Y$  表示林内  $\text{PM}_{2.5}$  质量浓度实测值;用最小二乘法计算求得常数项  $b_0$  及回归系数  $b_i$ ; $\varepsilon_i$  是回归误差,回归求解的过程就是使平均误差最小的过程。

通过逐步(stepwise)回归法将该研究全部的输入变量进行筛选,得到林内  $\text{PM}_{2.5}$  质量浓度  $\rho(\text{PM}_{2.5})$  的多元线性回归预测模型如下:

$$\rho(\text{PM}_{2.5}) = -19.2179 + 0.9689\rho(\text{PM}_{2.5}^0) + 0.0946H + 0.3425V_w - 0.4583T$$

预测模型显示,预测林内  $\text{PM}_{2.5}$  浓度时背景浓度的影响非常重要,其次是相对湿度。该预测模型经过逐步回归法后留取了 4 个输入变量(背景浓度、相对湿度、风速和温度),风向变量和季节变量未能进入回归。

### 3 结果与分析

**3.1 监测数据统计学描述** 在预测模型输入变量中,由于背景浓度(林外  $\text{PM}_{2.5}$  浓度)对林内  $\text{PM}_{2.5}$  浓度的预测结果影响最为重要,故将背景浓度数据与林内  $\text{PM}_{2.5}$  浓度数据分别描述,并进行  $T$  检验,以了解林内和林外  $\text{PM}_{2.5}$  浓度数据本身的差异,避免因对比数据无显著差异而不适合采用以上模型进行预测。由表 2 可知,林外  $\text{PM}_{2.5}$  浓度最小值高于林内,最大值低于林内, $\text{PM}_{2.5}$  小时平均浓度为林外高于林内; $T$  检验统计量观测值对应的双尾概率  $P$  值接近 0,如果设定显著水平  $\alpha$  为 0.05,则  $P$  值小于  $\alpha$ ,可认为林内和林外  $\text{PM}_{2.5}$  浓度存在显著差异。

**3.2 预测结果比较** 图 2 是基于 BP 人工神经网络方法和 MLR 方法仿真生成的研究区林内  $\text{PM}_{2.5}$  浓度的预测值与观测值的比对。图 2 显示 BP 人工神经网络模型与 MLR 模型

表 2 林内和林外监测数据描述及对比分析

Table 2 Description and comparative analysis of monitoring data inside and outside forests

项目 Item	μg/m <sup>3</sup>		
	林外 PM <sub>2.5</sub> 浓度 Concentration of PM <sub>2.5</sub> outside forest	林内 PM <sub>2.5</sub> 浓度 Concentration of PM <sub>2.5</sub> inside forest	林外 PM <sub>2.5</sub> 浓度-林内 PM <sub>2.5</sub> 浓度 Concentration of PM <sub>2.5</sub> outside forest-Concentration of PM <sub>2.5</sub> inside forest
极小值 Minimum	1.57	1.45	—
极大值 Maximum	512.12	542.27	—
均值 Mean	70.53	65.18	5.34
标准差 Standard deviation	79.33	76.03	13.07

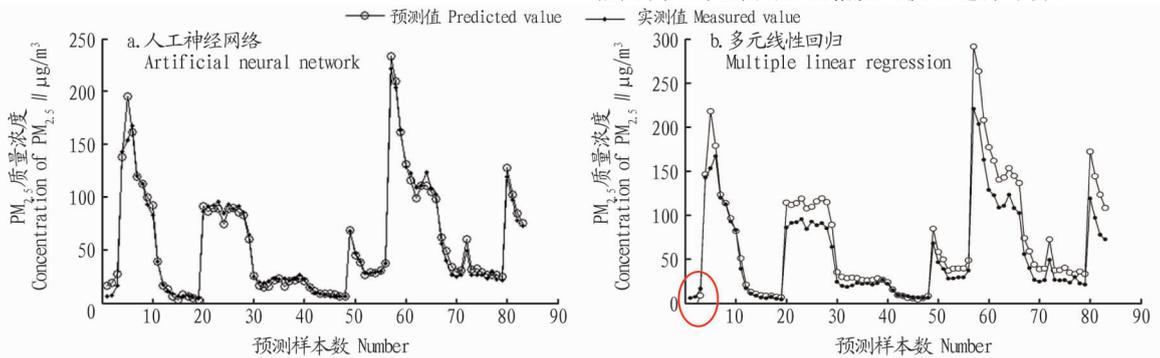
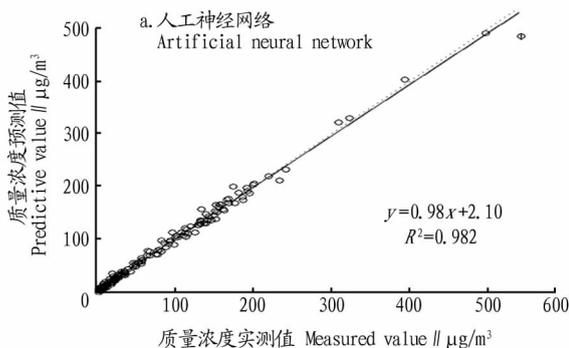


图 2 预测结果

Fig. 2 Prediction results

**3.3 预测精度评价** 可以选用多种预测方法对同一研究对象进行预测,但是预测效果往往存在不同程度的差异。预测精度即为评价预测效果优劣的标准,是衡量预测方法是否合理的依据。评价模型预测性能常用的检验指标<sup>[18-19]</sup>包括: $R^2$ ,表示拟合优度, $R^2$ 越接近 1,说明二者相关程度越高,预测结果与实测值越接近; $E_{MR}$ ,表示平均相对误差,预测的结果越精确, $E_{MR}$ 越小; $E_{RMS}$ ,表示均方根误差,是用来衡量预测值同实测值之间的偏差, $E_{RMS}$ 越小,预测结果越理想。以上 2 种预测方法的拟合优度见图 3,预测精度评价指标见表 3。

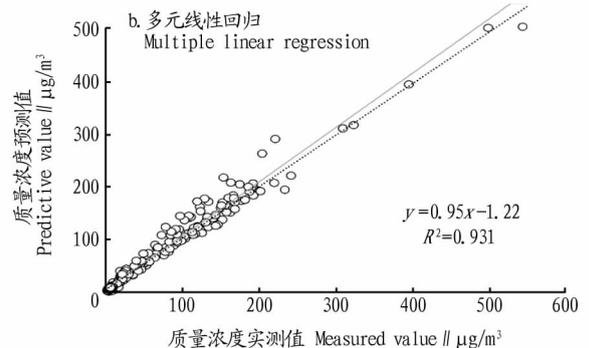
$$E_{MR} = (X-Y)/Y \quad (3)$$



$$E_{RMS} = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}} \quad (4)$$

式中, $X$ 为预测平均值; $Y$ 为实测平均值; $X_i$ 为第*i*组预测值; $Y_i$ 为第*i*组实测值; $n$ 为监测样本组数。

从图 3 可以看出,基于神经网络模型的拟合优度  $R^2$  为 0.982,而基于多元线性回归模型的拟合优度为 0.931,低于前者。从图 3a 和 b 拟合直线与期望值 1:1 直线关系也可以看出,神经网络模型的预测结果更加接近期望值,较多元线性回归模型预测结果的精确性有所提高。

图 3 PM<sub>2.5</sub> 小时浓度预测值与实测值相关关系Fig. 3 The correlation between predicted PM<sub>2.5</sub>-hour concentration and measured PM<sub>2.5</sub>-hour concentration

由表 3 可知,用神经网络预测的平均相对误差  $E_{MR}$  为  $1.71 \times 10^{-3}$ 、均方根误差  $E_{RMS}$  为 6.77;而 MLR 模型预测的平均相对误差  $E_{MR}$  为 0.27、均方根误差  $E_{RMS}$  为 22.92。虽然 2

种方法得到的预测值与实测值的相关性都很高(图 3),但从误差值大小可以看出,BP 神经网络模型预测的误差值远小于 MLR 模型。表 3 对比了 2 种模型预测精度,结果表明神

神经网络具有很强的稳定性和容错性<sup>[19]</sup>,综合信息能力强,能够很好地处理复杂的非线性关系。以上分析说明2种预测模型均能预测林内PM<sub>2.5</sub>浓度的小时平均值,相较而言,采用人工神经网络模型对林内PM<sub>2.5</sub>浓度值变化的预测结果比通过多元线性回归模型的预测结果更加接近实际观测值的变化。

表3 不同预测模型的预测精度

Table 3 Prediction accuracy of different prediction models

方法 Method	R <sup>2</sup>	E <sub>MR</sub>	E <sub>RMS</sub>
BP 神经网络 BP neural network	0.982	1.71×10 <sup>-3</sup>	6.77
MLR	0.931	0.27	22.92

#### 4 小结

(1)BP 神经网络和传统多元线性回归模型2种方法均可以用于林内大气PM<sub>2.5</sub>质量浓度预测,尽管二者在PM<sub>2.5</sub>浓度急剧变化时预测存在一定缺陷,但仍然能够在稀疏监测数据输入条件下基本反映林内PM<sub>2.5</sub>污染的时间变化规律。

(2)相对于多元线性回归模型,BP 神经网络具有很强的自学习、自组织与自适应功能,可以更加准确地实现林内PM<sub>2.5</sub>质量浓度小时平均值的预测,其预测结果能更好地捕捉气象因素对大气污染物浓度的非线性影响规律。BP 神经网络能够成为预测林内PM<sub>2.5</sub>质量浓度的较优方法。

(3)文中所选的污染源强表征变量和气象参数数据容易获得,且可以比较准确地反映气象变化和林分作用对PM<sub>2.5</sub>质量浓度的影响,是预测林内PM<sub>2.5</sub>质量浓度较为理想的输入变量。

(4)采用人工神经网络方法建立的林内PM<sub>2.5</sub>质量浓度预测模型适用于不同大气污染程度,但是在一般污染浓度情况下对林内PM<sub>2.5</sub>浓度的预测效果比高污染时期的预测效果更为准确。

#### 参考文献

[1] 邵龙义,时宗波,黄勤. 都市大气环境中可吸入颗粒物的研究[J]. 环境

保护,2000(1):24-26,29.

- [2] 孙淑萍,古润泽,张晶. 北京城区不同绿化覆盖率和绿地类型与空气中可吸入颗粒物(PM<sub>10</sub>) [J]. 中国园林,2004,20(3):77-79.
- [3] YU Y, SCHLEICHER N, NORRA S, et al. Dynamics and origin of PM<sub>2.5</sub> during a three-year sampling period in Beijing, China [J]. Journal of environmental monitoring, 2011, 13(2):334-346.
- [4] 吴海龙,余新晓,师忱,等. PM<sub>2.5</sub> 特征及森林植被对其调控研究进展 [J]. 中国水土保持科学,2012,10(6):116-122.
- [5] BECKETT K P, FREER-SMITH P H, TAYLOR G. Particulate pollution capture by urban trees: Effect of species and windspeed [J]. Global change biology, 2000, 6(8):995-1003.
- [6] LANGNER M. Reduction of airborne particulates by urban green [J]. Bundesamt für naturschutz, 2008, 179:129-137.
- [7] CAVANAGH J A E, ZAWAR-REZA P, WILSON J G. Spatial attenuation of ambient particulate matter air pollution within an urbanised native forest patch [J]. Urban forestry & urban greening, 2009, 8(1):21-30.
- [8] 郭二果,王成,郝光发,等. 北京西山典型游憩林空气悬浮颗粒物季节变化规律 [J]. 东北林业大学学报,2010,38(10):55-57.
- [9] 殷杉,蔡静萍,陈丽萍,等. 交通绿化带植物配置对空气颗粒物的净化效益 [J]. 生态学报,2007,27(11):4590-4595.
- [10] 汪永英,孙琪,李昭,等. 典型天气条件下哈尔滨城市森林不同林型对PM<sub>2.5</sub>的调控作用研究 [J]. 安徽农业科学,2016,44(5):175-179.
- [11] GARDNER M W, DORLING S R. Statistical surface ozone models: An improved methodology to account for non-linear behaviour [J]. Atmospheric environment, 2000, 34(1):21-34.
- [12] CHALOUKAKOU A, SAISANA M, SPYRELLIS N. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens [J]. Science of the total environment, 2003, 313(1/2/3):1-13.
- [13] MCKENDRY I G. Evaluation of artificial neural networks for fine particulate pollution (PM<sub>10</sub> and PM<sub>2.5</sub>) forecasting [J]. Journal of the air & waste management association, 2002, 52(9):1096-1101.
- [14] KOLEHMAINEN M, MARTIKAINEN H, RUUSKANEN J. Neural networks and periodic components used in air quality forecasting [J]. Atmospheric environment, 2001, 35(5):815-825.
- [15] PAPANASTASIOU D K, MELAS D, KIOUTSIUKIS I. Development and assessment of application of neural network and multiple regression models in order to predict PM<sub>10</sub> levels in a medium-sized Mediterranean city [J]. Water, air & soil pollution, 2007, 182:325-334.
- [16] 曹虹. 基于 BP 神经网络的交通流量预测 [D]. 西安:长安大学,2012:30.
- [17] 王敏,邹滨,郭宇,等. 基于 BP 神经网络的城市 PM<sub>2.5</sub> 浓度空间预测 [J]. 环境污染与防治,2013,35(9):63-70.
- [18] 石灵芝,邓红红,路婵,等. 基于 BP 神经网络的大气颗粒物 PM<sub>10</sub> 质量浓度预测 [J]. 中南大学学报(自然科学版),2012,43(5):1969-1974.
- [19] 宰松梅,郭冬冬,韩启彪,等. 基于人工神经网络理论的土壤水分预测研究 [J]. 中国农学通报,2011,27(8):280-283.

(上接第92页)

过低,繁殖系数降低,利于植株生长。在金钱草不定芽诱导培养、继代增殖培养及不定根诱导培养时,植物生长调节剂种类及浓度都在不断地变换,从而使金钱草茎能够快速增殖生长。在培养过程中发现,当添加相同浓度的 IBA 和 6-BA 时,主要诱导出金钱草不定根,抑制不定芽生长。IBA 浓度过高时,组培苗的叶片褶皱而卷曲,不利于植株生长。

#### 参考文献

[1] 《全国中草药汇编》编写组. 全国中草药汇编 [M]. 北京:人民卫生出版社,1986:538.

- [2] 江苏新医学院. 中药大词典(上册) [M]. 上海:上海科学技术出版社,1997:150-151.
- [3] 中国科学院西北植物研究所. 秦岭植物志:第1卷第4册 [M]. 北京:科学出版社,1983:29-54.
- [4] 李静,胡国全,桂敬强,等. 金钱草水提液对小鼠肾脏草酸钙结石形成的干预作用 [J]. 安徽科技学院学报,2014,28(2):5-11.
- [5] 李惠芝,姚崇舜,陈济民,等. 广金钱草粗提物中多糖的含量测定 [J]. 沈阳药科大学学报,1995,12(1):21-24.
- [6] 李静,贺绍君,刘德义. 金钱草防治泌尿系统结石机理研究进展 [J]. 辽宁中医药大学学报,2015,17(3):79-81.
- [7] 余伯阳. 中药生物技术 [M]. 北京:中国医药科技出版社,2005:195-196.
- [8] 梁定仁,黄明助,林伟,等. 植物生长调节剂对广金钱草扦插繁殖效果的影响 [J]. 南方农业学报,2016,47(4):614-617.