

薛冰,王启杰,马彬,等.基于MLR和神经网络的新疆杨日蒸腾量估算[J].华中农业大学学报,2023,42(5):240-250.  
DOI:10.13300/j.cnki.hnlkxb.2023.05.027

## 基于MLR和神经网络的新疆杨日蒸腾量估算

薛冰<sup>1</sup>,王启杰<sup>1</sup>,马彬<sup>2</sup>,梁雪<sup>1</sup>,侯振安<sup>1</sup>,姜艳<sup>1</sup>

1.石河子大学农学院,石河子832003;2.新疆伊犁哈萨克自治州特克斯县农业农村局种子站,伊犁835500

**摘要** 为解决当前旱区防护林主要树种新疆杨日蒸腾量( $T_r$ )估算值的精确度低、估算模型泛化能力差等问题,采用7种气象因子(日照时长、风速、相对湿度、饱和水蒸气压差、最低温、最高温和日均温)的8种组合作为模型输入,构建了传统多元线性回归模型(MLR)和神经网络模型(BP和Elman),估算2020年生长季新疆杨 $T_r$ 值,并对3种模型不同输入组合的估算结果进行比较与评价;同时基于Garson算法量化各气象因子对 $T_r$ 估算值的相对贡献率。结果显示,BP和Elman模型对 $T_r$ 估算值的精确度超过73.66%,在不同输入组合下神经网络模型估算精确度比MLR模型提高了8.45%~31.33%,其中拓扑结构为6-4-4-1的Elman模型估算值的精确度最高;气象因子饱和水蒸气压差对 $T_r$ 估算值的相对贡献率最大,相对湿度次之,不同温度变量对 $T_r$ 估算值的相对贡献率依次为:日均温>最低温>最高温。结果表明,所构建的新疆杨日蒸腾量的估算神经网络模型可提高对干旱地区防护林蒸腾量的估算精确度。

**关键词** 蒸腾量估算;多元线性回归;神经网络;新疆杨;防护林;水资源精准调控

**中图分类号** S727.23; TP183 **文献标识码** A **文章编号** 1000-2421(2023)05-0240-11

新疆杨(*Populus alba* var. *pyramidalis* Bunge)是杨柳科杨属植物<sup>[1]</sup>,是我国西北干旱区农田防护林系统的主要树种<sup>[2]</sup>,其蒸腾作用是涉及整个土壤-植被-大气连续体的复杂过程,但新疆杨日蒸腾量( $T_r$ )及其影响因素尚不明确,且传统的直接测量方法具有破坏性、适用范围有限等问题<sup>[3]</sup>。

近年来利用模型对植物蒸腾量进行间接估算成为研究热点。目前用于估算蒸腾量的模型包括SIM-DualKc、Shuttleworth-Wallace和多元线性回归(MLR)模型等<sup>[4-5]</sup>。然而,这些数学模型因局部参数化,需要大量样本观测数据导致其应用受到限制。更重要的是,植物蒸腾量受太阳辐射、温度、相对湿度、饱和水蒸气压差、风速等多种非线性气象因子的影响<sup>[6]</sup>,而传统数学模型无法完成非线性动态过程的精准模拟,因此,精确度较差。

机器学习模型中的人工神经网络(artificial neural network, ANN)可对输入与输出间繁复的关系进行映射,为非线性复杂系统过程的模拟和评判提供了有效途径,逐渐替代传统模型。其中误差反向传播神经网络(back propagation, BP)因在数据集不完

全可用的情况下,也能以高精度逼近任意非线性函数<sup>[7]</sup>,因此,在实际应用中80%~90%的人工神经网络都是采用BP神经网络或其变形<sup>[8]</sup>。Elman神经网络在结构上比BP神经网络多了1个承接层,增强了网络自身处理动态信息的能力<sup>[9-10]</sup>,更适用于非线性长期动态蒸腾量的估算。

目前神经网络在降水预报、径流预报、土壤含水量预测<sup>[11-12]</sup>等水文气象领域应用较为广泛,在植物蒸腾量模拟研究方面大多停留在农作物和经济林且多采用传统模型<sup>[13]</sup>,基于BP和Elman神经网络对农田防护林蒸腾耗水的研究相对较少<sup>[14]</sup>,尤其是在水资源紧缺的干旱地区。此外,目前已有的研究尚未就日最低(高)温和日均温作为输入变量对新疆杨日蒸腾量估算的单一和综合影响进行深入研究。因此,本研究采用BP和Elman神经网络建立适用于旱区新疆杨日蒸腾量估算的神经网络模型,并与MLR模型的估算结果进行比较,筛选出精确度更高的估算模型;同时明确各气象因子对BP和Elman神经网络模型估算 $T_r$ 值的相对贡献率,探讨气象因子的不同输入组合对神经网络模型估算干旱地区

收稿日期:2022-11-03

基金项目:国家自然科学基金项目(31660135);石河子大学科研项目(KX03100304)

薛冰, E-mail: xuebing@stu.sicau.edu.cn

通信作者:姜艳, E-mail: jiangyan098@163.com

植被蒸腾量的影响,以期防护林可持续发展和农业水资源精准调控提供科学指导。

## 1 材料与方法

### 1.1 研究区概况

本研究区位于新疆天山北麓准噶尔盆地西南缘,属于新疆生产建设兵团第八师150团(44°59'7"N, 86°8'56"E),该区紧邻古尔班通古特沙漠,属温带大陆性干旱荒漠气候。多年平均降水量为117.2 mm,潜在蒸发量可达1 942.1 mm。年均气温6.1℃,极端最高气温43.1℃,最低气温-42.8℃,热量资源丰富,风沙大<sup>[15]</sup>。

新疆杨(*Populus alba* var. *pyramidalis* Bunge)种植于2008年,株行距2 m×2 m,平均树高17 m,平均胸径14 cm,林区结构单一,林下无灌木,草本层主要以角果藜(*Ceratocarpus arenarius* L.)、沙蒿(*Artemisia desertorum* Spreng. Syst. Veg.)和叉毛蓬(*Petrosimonia sibirica* (Pall.) Bunge)等为主<sup>[16]</sup>。

### 1.2 气象要素与新疆杨蒸腾量监测

采用美国 Vantage Pro 2 自动气象站 (Davis Inc., USA) 每日观测最低温 ( $T_{\min}$ )、最高温 ( $T_{\max}$ )、日均温 ( $T_{\text{mean}}$ )、相对湿度 (relative humidity, RH)、日照时长 (SD)、风速 (U) 等气象数据,大气温度和湿度的协同效应采用饱和水蒸气压差 (vapor pressure difference, VPD) 表示。采用热扩散液流探针 (TDP30, Ecomatik, 德国) 观测树木液流,选取生长状况良好的样本树 6 株,在树干西侧距地面 160 cm 高处安装液流传感器。传感器由 2 根长 30 mm、直径 2 mm、相距 5 cm 的探针组成。两探针间的温差用数据采集器 (DL2e, Delta-T Devices, 英国) 自动记录。采用式

(1) 计算单株树干瞬时液流密度 ( $J$ , mL/(cm<sup>2</sup>·s)):

$$J = 0.714 \times \left( \frac{\Delta T_{\max} - \Delta T}{\Delta T} \right)^{1.231} \quad (1)$$

式(1)中, $\Delta T$ 为探针测定的瞬时温差,℃; $\Delta T_{\max}$ 为探针测定的昼夜最大温差,℃。

在林地中选择与样本树邻近且胸径相似的树木,用生长锥从胸高直径部位钻取木条,根据木条颜色深浅区分边材、心材直径,计算边材截面积:

$$S_A = \frac{\pi(D^2 - d^2)}{4} \quad (2)$$

式(2)中, $S_A$ 为边材截面积,cm<sup>2</sup>;  $D$ 为去皮直径,cm;  $d$ 为髓心直径,cm。

新疆杨单株日蒸腾量 ( $T_r$ , mL/d) 计算公式如下:

$$T_r = \frac{\sum_j^M (J_j \times S_{A,j} \times 3600 \times 24)}{M} \quad (3)$$

式(3)中, $M$ 为样本树的个数; $J_j$ 和 $S_{A,j}$ 分别为第 $j$ 棵样本树的树干瞬时液流密度(mL/(cm<sup>2</sup>·s))和边材截面积(cm<sup>2</sup>)。

### 1.3 样本选择

为了探究气象因子对  $T_r$  估算值的影响,本研究在2020年5月—2021年4月期间观测记录气象因子和  $T_r$  值,共收集数据 365 组。以新疆杨生长季(2020年5月—2020年10月)气象数据作为输入变量,输出变量则采用实测  $T_r$  值。本研究首先采用逐步回归方法筛选出 SD、U、RH、VPD、 $T_{\min}$ 、 $T_{\max}$ 、 $T_{\text{mean}}$  7 个气象因子作为模型输入变量,其次,为重点评估温度变量对  $T_r$  估算值的影响,将  $T_{\min}$ 、 $T_{\max}$ 、 $T_{\text{mean}}$  与 SD、U、RH、VPD 进行组合,最终将 7 种变量分成 8 种输入组合(表1)。

表1 多元线性回归(MLR)、BP神经网络和 Elman神经网络模型气象变量的输入组合

Table 1 Input combinations of meteorological variables used in multiple linear regression (MLR), BP neural network and Elman neural network models

序号 Number	模型 Models			输入组合 Input combinations						
1	MLR1	BP1	Elman1	SD	U	RH	VPD			
2	MLR2	BP2	Elman2	SD	U	RH	VPD	$T_{\min}$		
3	MLR3	BP3	Elman3	SD	U	RH	VPD	$T_{\max}$		
4	MLR4	BP4	Elman4	SD	U	RH	VPD	$T_{\text{mean}}$		
5		BP5	Elman5	SD	U	RH	VPD	$T_{\min}$	$T_{\max}$	
6		BP6	Elman6	SD	U	RH	VPD	$T_{\min}$	$T_{\text{mean}}$	
7		BP7	Elman7	SD	U	RH	VPD	$T_{\max}$	$T_{\text{mean}}$	
8		BP8	Elman8	SD	U	RH	VPD	$T_{\min}$	$T_{\max}$	$T_{\text{mean}}$

将新疆杨生长季数据集随机分为2个子集,其中80%用于模型训练(148组),其余20%用于模型测试(36组),训练和测试过程中输入与输出变量的统计结果如表2所示,发现在训练和测试阶段数值的差

异很小。由于各指标量纲或数量级存在差异,在使用80%样本数据进行模型建立之前,为更好地反映各因素之间的相互作用,防止部分神经元达到过饱和和状态,对所有样本数据进行了归一化处理。

表2 训练和测试过程中输入与输出变量的统计

Table 2 Statistics of input and output variables during training and testing

统计 Statistics		SD/h	U/(m/s)	RH/%	VPD/kPa	$T_{\min}/^{\circ}\text{C}$	$T_{\max}/^{\circ}\text{C}$	$T_{\text{mean}}/^{\circ}\text{C}$	$T_r/(\text{mL}/\text{d})$
训练集 Training set	最小值 Minimum	0.00	0.50	12.50	0.35	-1.00	11.80	7.30	148.50
	平均值 Mean	9.86	1.90	41.23	1.61	14.57	29.01	21.64	722.90
	最大值 Maximum	14.70	5.20	78.80	3.36	23.30	37.30	30.70	2 183.88
	标准差 Standard deviation	3.67	0.74	12.64	0.63	5.05	5.94	5.40	397.95
测试集 Test set	最小值 Minimum	4.70	0.70	16.30	0.70	-1.40	10.50	4.00	114.70
	平均值 Mean	10.46	1.63	45.85	1.63	10.23	25.04	17.42	837.39
	最大值 Maximum	14.50	3.00	66.50	3.00	18.70	36.90	28.20	1 966.14
	标准差 Standard deviation	2.43	0.51	12.21	0.51	7.28	8.08	7.88	436.86

#### 1.4 多元线性回归模型(MLR)

MLR用于对多个自变量和一个或多个因变量之间的关系进行建模,本研究基于气象因子和 $T_r$ 、以及不同输入变量之间的关系,建立MLR模型如下:

$$T_r = a_0 + a_1x_1 + a_2x_2 + \dots + a_{k-1}x_{k-1} + a_kx_k \quad (4)$$

式(4)中, $x$ 为输入变量(气象因子); $a$ 为各输入变量所对应的偏回归系数; $k$ 为输入变量的个数。

#### 1.5 BP和Elman神经网络模型

BP人工神经网络是采用误差反向传播算法的静态前馈网络,通常由输入层、隐含层和输出层构成,Elman神经网络在结构上比BP网络多了1个承接层,增加了网络自身处理时间序列信息的能力(图1)。因为存在8种输入组合,所以BP和Elman神经网络模型的输入层节点数依次为4、5、5、5、6、6和7;输出层的节点数与模型输出的因变量个数相同,即1;隐含层节点数的确定,先采用10折交叉验证法对数据集进行多次划分,再对多次评估结果取平均值,最后根据最小均方根误差准则确定,同时消除了单次划分时数据划分不平衡而造成的不良影响,在Elman神经网络中承接层节点数等于隐藏层节点数。最终依次建立拓扑结构为4-4-1、5-3-1、5-4-1、5-6-1、6-3-1、6-3-1、6-4-1和7-5-1的BP神经网络蒸腾模型,以及拓扑结构为4-8-8-1、5-4-4-1、5-6-6-1、5-5-5-1、6-7-7-1、6-4-4-1、6-10-10-1、7-9-9-1的Elman神经网络蒸腾模型。

本研究在MATLAB R2016a环境下对BP和Elman神经网络模型进行构建,输入层与隐含层之间

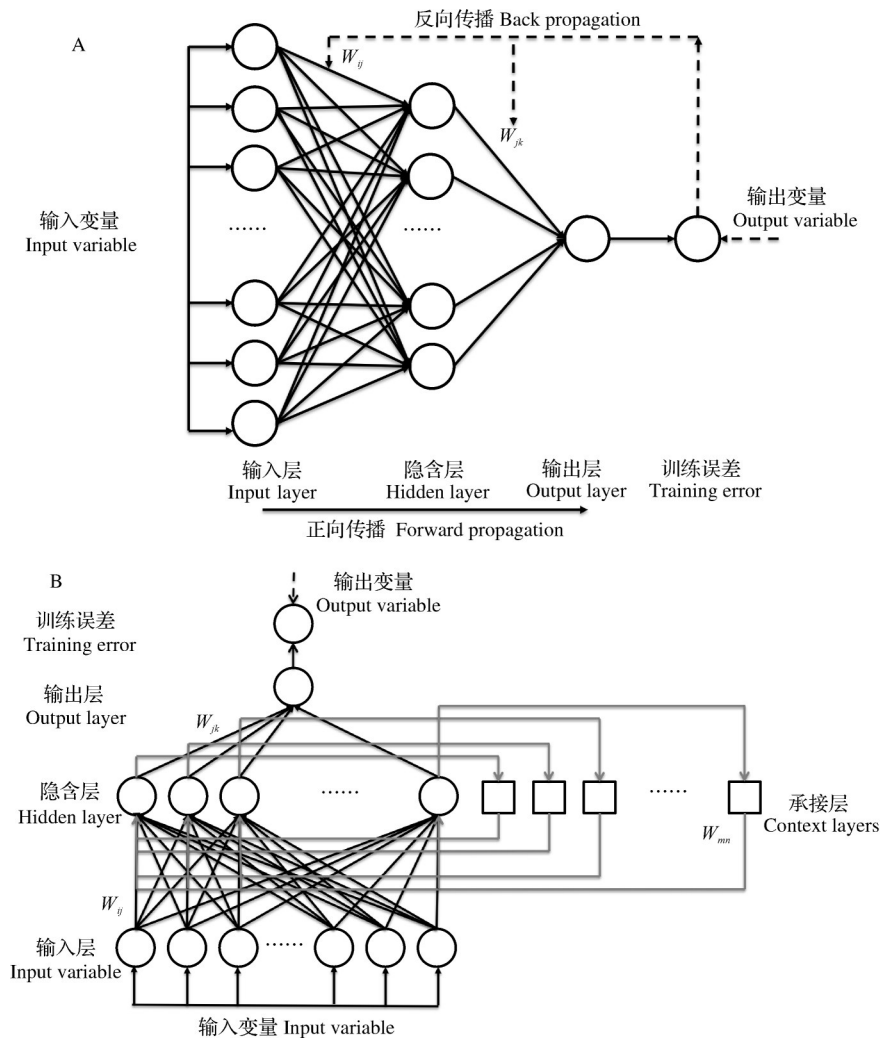
的传递函数设为正切S型函数tansig,隐含层与输入层之间的传递函数为线性函数purelin,采用Levenberg-Marquardt(LM)优化算法(trainlm)作为网络的训练函数,设置训练目标最小误差为0.000 1,动态因子为0.01,最大迭代次数为1 000,最小性能梯度为 $1 \times 10^{-6}$ ,最高失败次数为6。

#### 1.6 模型评价

随机选取20%的样本数据用于网络性能测试,使用5种统计参数量化MLR、BP和Elman模型模拟结果的偏差,分别为决定系数( $R$ -square,  $R^2$ )、纳什效率系数(Nash sutcliffe efficiency coefficient, NSE)、平均绝对误差(mean absolute error, MAE)、均方根误差(root mean squared error, RMSE)和归一化均方根误差(normalized root mean squared error, NRMSE),计算方法参考文献[13]。模型的估算精确度(accuracy,  $A$ )用 $1 - \text{MAPE}$ (mean absolute percentage error, 平均绝对百分比误差)表示。 $R^2$ 和NSE越接近1,并且MAE、RMSE和NRMSE的值越接近0,模型估算值就越准确。此外,NRMSE还具有模型精确度的定性作用,若NRMSE值大于0.3,则认为模型的性能较差;如果其值在0.2~0.3,则认为模型的性能中等;如果其值在0.1~0.2,则认为模型的性能很好;若其值小于0.1,则认为模型的性能极好<sup>[17]</sup>。

$$A = 1 - \text{MAPE} = 1 - \frac{100\%}{N} \sum_{i=1}^N \left| \frac{O_i - P_i}{O_i} \right| \quad (5)$$

式(5)中, $N$ 为样本数; $P_i$ 和 $O_i$ 分别是第 $i$ 估算值和实测值。



$W_{ij}$ 、 $W_{jk}$  和  $W_{mn}$  分别为输入层到隐含层、隐含层到输出层和承接层到隐含层的连接权重。 $W_{ij}$ 、 $W_{jk}$  和  $W_{mn}$  are the connection weights from input layer to hidden layer, hidden layer to output layer and undertake layer to hidden layer, respectively.

图1 BP(A) 和 Elman 人工神经网络拓扑结构(B)

Fig. 1 Topological structure of BP(A) and Elman artificial neural network(B)

### 1.7 ANN模型中各输入变量对输出变量的相对贡献率

采用 Garson<sup>[17]</sup>算法,利用连接权重分析神经网络各输入变量对输出变量的相对贡献率,计算公式如下:

$$R_{ij} = \frac{\sum_{j=1}^L \left( |W_{ij}W_{jk}| / \sum_{r=1}^N |W_{rj}| \right)}{\sum_{i=1}^N \sum_{j=1}^L \left( |W_{ij}W_{jk}| / \sum_{r=1}^N |W_{rj}| \right)} \times 100\% \quad (6)$$

式(6)中, $R_{ij}$ 为输入变量对输出变量的相对贡献率,%; $i$ 、 $j$ 和 $k$ 分别为输入层、隐含层和输出层的节点数; $W_{ij}$ 、 $W_{jk}$ 分别为输入层-隐含层、隐含层-输出层的连接权值; $N$ 和 $L$ 分别为输入层和隐含层的节点总数。

### 1.8 数据处理

采用SPSS 26.0对数据进行统计分析;在MATLAB R2016a环境下构建BP和Elman神经网络模型,并基于SPSS完成逐步回归和MLR模型;采用MATLAB绘制泰勒图,Origin 2021绘制其他图。

## 2 结果与分析

### 2.1 气象因子与新疆杨蒸腾量的时间变化特征

由表3可知,在观测期内新疆杨蒸腾量与气象因子有明显的季节性变化,新疆杨单株全年蒸腾量为  $1.5950 \times 10^5$  mL,生长季单株蒸腾量为  $1.3839 \times 10^5$  mL,非生长季单株蒸腾量为  $2.1109 \times 10^4$  mL,生长季蒸腾量占全年蒸腾量的86.77%,单株月均蒸腾量  $1.3291 \times 10^4$  mL。在生长季6月新疆杨蒸腾量最大,

表3 观测期内气象因子与新疆杨日蒸腾量的趋势变化

Table 3 Trends of meteorological factors and daily transpiration of *Populus alba* var. *pyramidalis* Bunge during the observation period

统计 Statistics		SD/h	U/(m/s)	RH/%	VPD/kPa	$T_{\min}/^{\circ}\text{C}$	$T_{\max}/^{\circ}\text{C}$	$T_{\text{mean}}/^{\circ}\text{C}$	$T_r/(\text{mL}/\text{d})$
1月 Jan	最小值 Minimum	0.10	0.30	71.30	0.02	-30.90	-22.80	-27.60	37.23
	平均值 Mean	5.83	0.90	77.67	0.04	-21.60	-13.40	-17.92	43.23
	最大值 Maximum	9.80	1.30	84.00	0.09	-14.30	-1.30	-8.40	53.50
2月 Feb	最小值 Minimum	0.00	0.70	42.30	0.02	-32.20	-22.90	-27.60	41.61
	平均值 Mean	4.81	1.35	73.13	0.11	-12.73	-3.23	-8.09	61.82
	最大值 Maximum	10.30	3.30	92.80	0.52	1.10	12.60	5.40	86.42
3月 Mar	最小值 Minimum	0.00	0.60	12.00	0.08	-12.80	-2.40	-7.90	29.94
	平均值 Mean	6.64	1.36	57.23	0.38	-2.41	9.19	3.07	99.34
	最大值 Maximum	12.60	2.80	85.80	1.41	6.70	22.50	14.00	147.93
4月 Apr	最小值 Minimum	0.00	0.70	12.00	0.07	-4.80	0.30	-0.60	21.53
	平均值 Mean	7.99	1.94	44.35	0.72	3.01	15.46	9.02	84.78
	最大值 Maximum	12.60	3.50	88.30	1.52	11.90	24.10	16.60	118.52
5月 May	最小值 Minimum	0.00	0.50	12.50	0.36	0.60	16.70	8.90	291.87
	平均值 Mean	9.44	2.01	42.30	1.23	11.48	25.21	18.22	502.40
	最大值 Maximum	14.70	5.20	78.80	1.89	18.60	33.00	23.80	916.10
6月 June	最小值 Minimum	3.70	0.90	15.30	0.93	11.80	24.90	18.60	291.67
	平均值 Mean	11.36	1.92	35.22	2.00	16.55	31.49	24.10	1 168.81
	最大值 Maximum	14.60	3.50	63.50	3.36	23.30	37.30	30.70	1 966.14
7月 July	最小值 Minimum	2.80	1.10	16.30	1.00	13.40	28.20	19.20	275.61
	平均值 Mean	11.48	2.05	39.57	1.93	16.89	32.73	24.85	1 018.19
	最大值 Maximum	14.60	3.30	63.70	2.76	21.50	37.30	28.70	2 183.88
8月 Aug	最小值 Minimum	2.60	1.20	14.50	0.74	11.80	24.30	18.20	309.83
	平均值 Mean	9.91	1.90	43.76	1.84	18.19	32.75	25.23	764.04
	最大值 Maximum	14.70	3.50	66.80	3.36	22.50	36.20	28.70	1 518.83
9月 Sept	最小值 Minimum	1.30	0.60	31.30	0.62	5.60	16.50	10.80	240.94
	平均值 Mean	8.71	1.64	46.54	1.25	12.54	26.83	19.38	669.18
	最大值 Maximum	13.90	2.90	66.50	2.08	19.00	36.90	28.00	942.52
10月 Oct	最小值 Minimum	0.00	0.60	23.80	0.35	-1.40	10.50	4.00	114.70
	平均值 Mean	8.96	1.59	46.05	0.99	7.17	20.65	13.52	400.83
	最大值 Maximum	14.60	4.80	67.50	2.31	18.10	32.60	24.90	811.17
11月 Nov	最小值 Minimum	0.00	0.60	18.30	0.03	-15.60	-6.70	-11.20	176.91
	平均值 Mean	8.71	1.58	49.99	0.77	3.80	17.30	10.24	388.48
	最大值 Maximum	12.10	3.00	87.00	1.95	18.40	28.00	22.40	557.35
12月 Dec	最小值 Minimum	0.00	0.20	68.80	0.02	-29.60	-21.70	-26.70	20.66
	平均值 Mean	3.50	1.05	78.71	0.04	-19.93	-12.01	-15.95	27.25
	最大值 Maximum	7.10	2.30	84.50	0.07	-8.70	-1.60	-7.90	35.97

为  $3.5064 \times 10^4$  mL, 其次为7月、8月、9月、5月和10月, 单株月蒸腾量分别为  $3.1564 \times 10^4$ 、 $2.3685 \times 10^4$ 、 $2.0075 \times 10^4$ 、 $1.5574 \times 10^4$  和  $1.2426 \times 10^4$  mL。在非生长季11月新疆杨蒸腾量最大, 为  $1.1654 \times 10^4$  mL, 其次为3月、4月、2月、1月和12月, 单株月蒸腾量分别为  $3.0795 \times 10^3$ 、 $2.4588 \times 10^3$ 、 $1.7309 \times 10^3$ 、 $1.3402 \times$

$10^3$  和  $8.4490 \times 10^2$  mL。

气象因子SD、U、VPD、 $T_{\min}$ 、 $T_{\max}$  和  $T_{\text{mean}}$  与  $T_r$  呈现了相似的季节性变化,  $T_{\max}$  为  $37.3^{\circ}\text{C}$ , 出现在6月、7月,  $T_{\min}$  为  $-22.8^{\circ}\text{C}$ , 出现在1月; 生长季SD明显比非生长季更长; U最大值出现在春秋两季; 而RH在非生长季比生长季大, 这可能是由于非生长季蒸腾

散发减小,气温下降,同时降雪补给所造成的。

为了衡量各因子的变异程度,采用雷达图分析观测期内各月气象因子与 $T_r$ 实测值的变异系数(图2),可以看出,SD和U在生长季的变异程度比非生长季的小(图2A);RH、VPD、 $T_{min}$ 、 $T_{max}$ 和 $T_{mean}$ 在春秋两季的变异程度比冬夏两季的变异程度大, $T_{min}$ 、 $T_{max}$ 和 $T_{mean}$ 在生长季中除10月以外,其他月份的变异系数均小于0.5,而在非生长季存在变异系数大于1的现象(图2B、2C)。但对于 $T_r$ 值,生长季的变异系数明显比非生长季的变异系数大(图2D)。

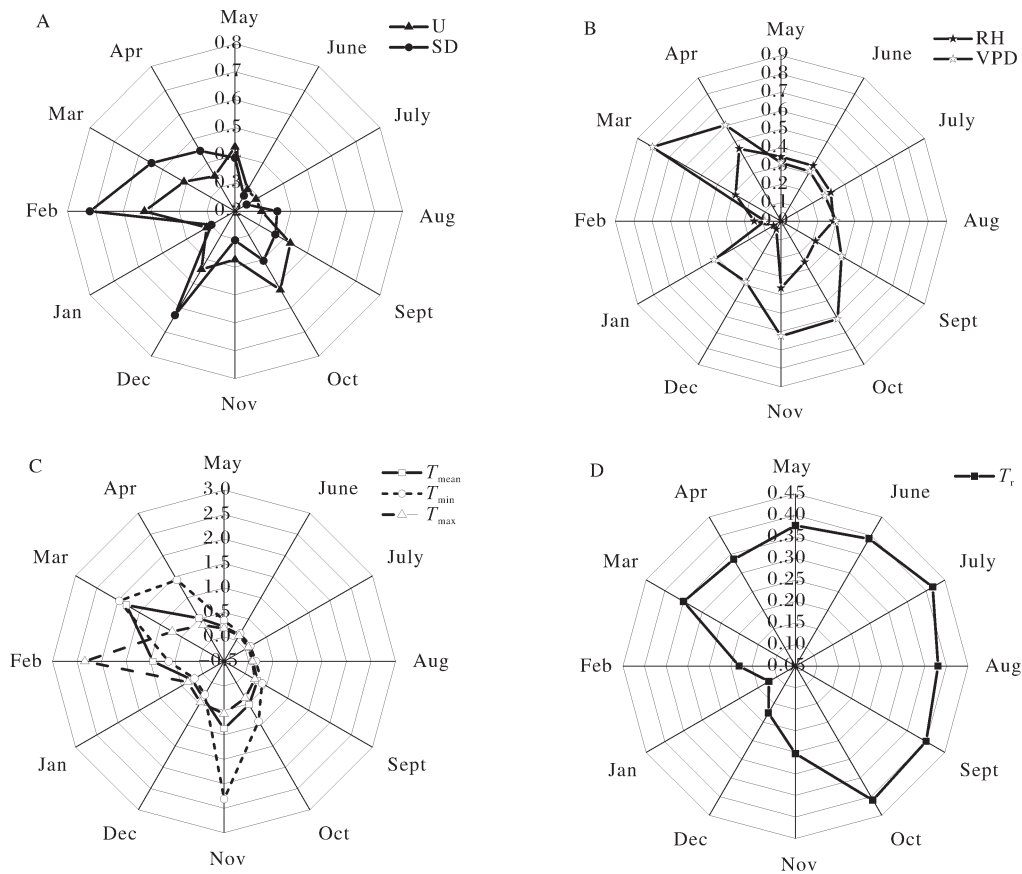
### 2.2 不同输入组合对 $T_r$ 估算值的影响

表4为MLR、BP和Elman模型在不同输入组合下 $T_r$ 值的估算结果,可以看出随着输入变量的改变,BP和Elman模型估算性能相对稳定,其中,Elman神经网络模型在温度变量为 $T_{min}$ 和 $T_{mean}$ 时表现最好,输入变量包括所有7个气象因素时表现最差;BP模型在输入变量包括所有7个气象因素时表现最好,而

在温度变量为 $T_{max}$ 时表现最差。传统MLR模型输入变量为不同类型温度时,模型的估算精确度明显不同,当 $T_{min}$ 、 $T_{max}$ 或 $T_{mean}$ 作为模型输入变量比未考虑温度的MLR模型拟合精确度分别提高13.74%、23.37%、32.09%;以RMSE为标准,估算偏差分别降低41.14%、36.87%、72.34%。

### 2.3 最佳估算模型的选择

如表4所示, $T_r$ 估算值在不同模型之间也存在一定差异。其中Elman模型精确度(73.66%~82.88%)最高,相对于MLR模型, $R^2$ 和NSE分别提高了2.29%~57.03%、1.82%~57.42%;MAE、RMSE和NRMSE分别降低了0.12%~77.99%、5.62%~89.23%和5.68%~89.28%。BP模型精确度(75.05%~80.52%)略低于Elman模型, $R^2$ 为0.86~0.88。总体而言,当输入变量相同时,与线性MLR模型相比,2种非线性ANN模型均能够以较高准确度估算 $T_r$ 值。与缺少温度的MLR模型相比,具有 $T_{min}$ 、 $T_{max}$ 或 $T_{mean}$ 输入变量的MLR模型显示出较高的拟合度。



A: U, SD; B: RH, VPD; C:  $T_{mean}$ ,  $T_{max}$ ,  $T_{min}$ ; D:  $T_r$ .

图2 观测期内每月不同气象因子与新疆杨日蒸腾量的变异系数

Fig. 2 Coefficient of variation of monthly different meteorological factors and daily transpiration of *Populus alba* var. *pyramidalis* Bunge during the observation period

表4 多元线性回归(MLR)、BP神经网络和Elman神经网络模型在各种输入组合下的统计值  
 Table 4 Statistical values of multiple linear regression (MLR), BP neural network and Elman neural network models under various input combinations

序号 Number	模型 Models	统计 Statistics	输入参数 Input parameters	R <sup>2</sup>	NSE	MAE/ (mL/d)	RMSE/ (mL/d)	NRMSE	A/%
1	MLR	$T_r = 36.226X_{SD} + 127.039X_U - 0.813X_{RH} - 343.147X_{VPD} - 127.461$	SD, U, RH, VPD	0.56	0.56	220.80	290.91	0.16	42.33
2		$T_r = 39.501X_{SD} + 102.851X_U - 6.469X_{RH} + 151.057X_{VPD} + 15.891T_{min} - 208.516$	SD, U, RH, VPD, T <sub>min</sub>	0.80	0.78	162.57	206.11	0.11	56.08
3		$T_r = 33.308X_{SD} + 122.808X_U - 4.100X_{RH} + 217.277X_{VPD} + 10.496T_{max} - 34.765$	SD, U, RH, VPD, T <sub>max</sub>	0.77	0.76	167.09	212.54	0.11	65.71
4	BP	$T_r = 35.027X_{SD} + 105.512X_U - 6.717X_{RH} + 122.072X_{VPD} + 17.668T_{mean} - 150.679$	SD, U, RH, VPD, T <sub>mean</sub>	<b>0.86</b>	<b>0.85</b>	<b>131.25</b>	<b>168.80</b>	<b>0.09</b>	<b>74.43</b>
5		4-4-1	SD, U, RH, VPD	0.87	0.86	127.69	161.41	0.09	75.05
6		5-3-1	SD, U, RH, VPD, T <sub>min</sub>	0.87	0.86	127.62	163.17	0.09	69.86
7	拓扑结构 Topological structure	5-4-1	SD, U, RH, VPD, T <sub>max</sub>	0.87	0.85	123.90	167.21	0.09	79.57
8		5-6-1	SD, U, RH, VPD, T <sub>mean</sub>	0.87	0.87	121.63	159.06	0.09	76.25
9		6-3-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>max</sub> , T <sub>mean</sub>	0.88	0.86	126.54	164.57	0.09	80.52
10	Elman	6-3-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>mean</sub>	0.88	0.88	120.93	153.50	0.08	69.70
11		6-4-1	SD, U, RH, VPD, T <sub>max</sub> , T <sub>mean</sub>	0.87	0.87	127.50	156.33	0.08	76.38
12		7-5-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>max</sub> , T <sub>mean</sub>	<b>0.88</b>	<b>0.88</b>	<b>118.62</b>	<b>152.38</b>	<b>0.08</b>	<b>78.38</b>
13	Elman	4-8-8-1	SD, U, RH, VPD	0.88	0.88	124.05	153.73	0.08	73.66
14		5-4-4-1	SD, U, RH, VPD, T <sub>min</sub>	0.88	0.87	120.21	156.59	0.08	80.42
15		5-6-6-1	SD, U, RH, VPD, T <sub>max</sub>	0.87	0.87	118.31	159.99	0.09	80.59
16	Elman	5-5-5-1	SD, U, RH, VPD, T <sub>mean</sub>	0.87	0.87	131.10	159.82	0.09	75.28
17		6-7-7-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>max</sub>	0.88	0.86	109.84	161.29	0.09	82.42
18		6-4-4-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>mean</sub>	<b>0.89</b>	<b>0.89</b>	<b>108.58</b>	<b>146.05</b>	<b>0.08</b>	<b>82.88</b>
19	Elman	6-10-10-1	SD, U, RH, VPD, T <sub>max</sub> , T <sub>mean</sub>	0.88	0.87	122.98	159.77	0.09	78.28
20		7-9-9-1	SD, U, RH, VPD, T <sub>min</sub> , T <sub>max</sub> , T <sub>mean</sub>	0.87	0.86	128.46	162.55	0.09	78.31

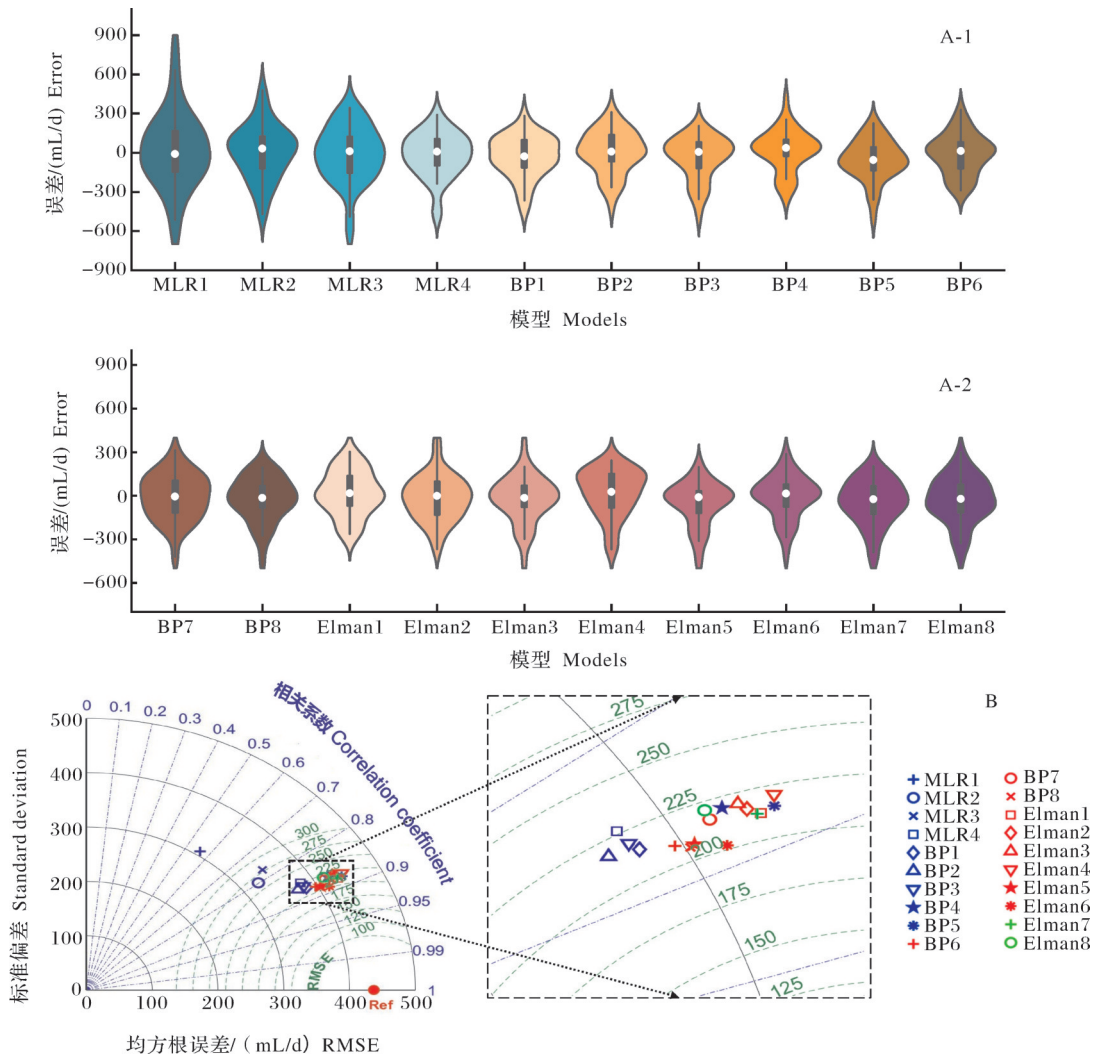
注:3种模型中最高的统计指标用粗体标出。Note: The best statistical indicators of the 3 models were marked in bold.

为了进一步探索不同模型  $T_r$  实测值和估算值的差异,研究在测试阶段的不同输入组合下,通过小提琴图分析MLR、BP和Elman模型  $T_r$  实测值和估算值的误差分布(图3A)。小提琴图结合了箱形图和密度图的特征,可以形象地展示多组数据的分布状态以及概率密度。可以看出,在各种输入组合下ANN模型实测值和估算值的误差分布明显较小,而MLR模型实测和估算值误差分布较大,其中前3个输入组合下误差分布最大,且有高估  $T_r$  值的现象。图3B使用泰勒图将3种模型的不同输入组合对估算  $T_r$  值的整体性能进行评估,所有结果均证实BP和Elman神经网络模型的性能高于传统的MLR模型,可用于估算  $T_r$  值。拓扑结构为6-4-4-1的Elman模型性能最高,可作为估算  $T_r$  值的优选模型。但发现随着输

入变量的增加,BP和Elman估算  $T_r$  值的性能差异不明显,甚至有BP估算性能比Elman更好的趋势。

### 2.4 气象因子对ANN模型估算值的相对贡献率

BP和Elman神经网络在8种不同输入组合下,各气象因子对  $T_r$  估算值的贡献率不同(表5)。在大多数输入组合下,最大相对贡献率由VPD产出,其次为RH,两者的相对贡献率之和位于30.68%~59.42%。BP(7-5-1)和Elman(6-4-4-1)模型在最佳输入组合下对  $T_r$  估算值产生最大贡献的输入变量分别为RH(19.11%)和VPD(22.03%)。3种温度变量在2种ANN模型中贡献率变化表现出相同趋势,当输入组合存在2种温度变量时,若其中1个变量为  $T_{mean}$ ,则  $T_{mean}$  的相对贡献率比另一温度变量( $T_{min}$ 或



A: 小提琴图; B: 泰勒图。A: Violin plot; B: Taylor diagram.

图3 MLR、BP和Elman神经网络模型在不同输入组合下  $T_r$  估算值的性能评价

Fig.3 Performance evaluation of MLR, BP and Elman neural network models for estimating  $T_r$  values under various input combinations



$T_{\max}$ )的相对贡献率大;当 $T_{\min}$ 和 $T_{\max}$ 同时作为输入变量时, $T_{\min}$ 的相对贡献率比 $T_{\max}$ 的贡献率大,即 $T_{\text{mean}} > T_{\min} > T_{\max}$ ,在MLR模型中呈现相同趋势。当存在温度变量时,其他气象因子(SD、U、RH和VPD)贡献率之和在60.90%~85.68%(BP)、61.80%~86.36%(Elman)。

表5 BP和Elman神经网络模型在不同组合下各输入参数的相对贡献率

**Table 5 Relative contribution rate of each input variable to the output variable under the eight input combinations of BP and Elman neural network model** %

模型 Models	拓扑结构 Topological structure	SD	U	RH	VPD	$T_{\min}$	$T_{\max}$	$T_{\text{mean}}$
BP1	4-4-1	29.50	13.23	42.67	14.60			
BP2	5-3-1	13.90	12.23	26.34	32.65	14.88		
BP3	5-4-1	10.89	11.74	30.07	29.35		17.95	
BP4	5-6-1	24.09	18.17	21.02	22.40			14.32
BP5	6-3-1	23.77	11.39	6.87	23.81	17.79	16.37	
BP6	6-3-1	21.01	6.32	7.30	29.62	14.83		20.92
BP7	6-4-1	14.11	8.32	16.23	22.24		17.74	21.36
BP8	7-5-1	14.88	14.20	19.11	18.33	8.93	11.46	13.09
Elman1	4-8-8-1	26.24	21.16	34.92	17.68			
Elman2	5-4-4-1	11.24	10.39	27.80	27.84	22.73		
Elman3	5-6-6-1	12.65	17.89	27.67	28.15		13.64	
Elman4	5-5-5-1	25.85	14.06	12.65	20.14			27.3
Elman5	6-7-7-1	15.78	11.61	20.60	18.75	20.47	12.79	
Elman6	6-4-4-1	18.15	10.81	10.81	22.03	18.78		19.42
Elman7	6-10-10-1	11.56	13.14	17.23	23.26		13.57	21.24
Elman8	7-9-9-1	16.76	17.28	17.86	16.41	12.01	7.63	12.05

### 3 讨论

为准确估算新疆杨日蒸腾量,并明确各气象因子对蒸腾量的贡献,本研究利用SD、U、RH、VPD、 $T_{\min}$ 、 $T_{\max}$ 、 $T_{\text{mean}}$  7种气象因子的8种组合作为模型输入,建立了用于新疆杨日蒸腾量估算的ANN神经网络模型(BP和Elman)和传统MLR模型,通过 $R^2$ 、NSE、MAE、RMSE和NRMSE这5种统计参数对MLR、BP和Elman模型在不同输入变量下的性能进行了评价。结果表明,随着输入变量的增加,2种ANN模型估算 $T_r$ 值的精确度变化不大,这与杜斌等<sup>[18]</sup>通过监测玉米液流得出的结论存在差异,原因可能是由于不同植物类型对应模型参数的敏感性不同。总而言之,BP和Elman模型在各输入组合下均能以高精度(73.66%~82.88%)估算 $T_r$ 值,有效表征气象因子与 $T_r$ 之间的内在联系;其中拓扑结构为6-4-4-1的Elman模型估算精确度(82.88%)最高,可作为新疆杨蒸腾耗水量估算的优选模型。

本研究利用Garson算法分析BP和Elman模型

中各气象因子对 $T_r$ 估算值的相对贡献率,发现在2种ANN模型中,RH和VPD是 $T_r$ 估算值的主要影响因素(表5),与Aghelpour等<sup>[13]</sup>和贾国栋等<sup>[19]</sup>的研究结论一致。ANN神经网络模型的估算精确度和稳定性明显高于传统MLR模型,整体性能更好(图3B);对于传统MLR模型,研究表明在确保输入变量关联度尽可能低的情况下,输入变量应包括 $T_{\min}$ 、 $T_{\max}$ 和 $T_{\text{mean}}$ 中的一种(其中 $T_{\text{mean}}$ 最佳, $T_{\min}$ 次之),否则 $T_r$ 估算值可能会出现较大误差(表4)。此外,MLR模型的计算结果在一定程度上高估了 $T_r$ 值(图3A)。

利用人工神经网络模型估算新疆杨日蒸腾量具有明显优势,可能是因为神经网络从信息处理的角度抽象人脑神经元并建立模型,通过训练使网络的 $T_r$ 估算值与实测值趋近,精准反映了气象因子与 $T_r$ 的非线性关系。但在ANN模型中,隐藏层神经元个数显著影响神经网络模型的模拟精确度,隐含层神经元太少会损害网络的功能,而隐含层神经元过多又会导致网络过度拟合数据<sup>[20]</sup>。基于此,本研究采用10折交叉验证的方法对模型调优,得到使模型泛

化性能最优的隐含层神经元个数,然后利用训练集重新训练模型,并使用独立测试集对模型性能做出评价,弥补了数据集单次划分造成的估算偏差,从而达到了精准估算未来时段蒸腾量的目的,可为旱区新疆杨生长季蒸腾量预报和农业灌溉管理提供更精准的科学指导。

蒸腾反映了植物水分状况对环境变化的适应策略,土壤水热状况、气象因素等均会对蒸腾量产生影响<sup>[21]</sup>,新疆非生长季气温极低,林木蒸腾作用大幅降低,基本处于休眠状态,非生长季蒸腾量仅占全年蒸腾总量的13.23%(表3)。因此,生长季气象及蒸腾量数据是本研究建模的重点,但是否适用于非生长季有待进一步深入分析,也是未来的研究重点。

## 参考文献 References

- [1] ZHANG S G. Chromosome numbers of some *Populus* taxa from China[J]. *Acta phytotaxonomica sinica*, 2005, 43(6): 539-544.
- [2] WU N, LI Z, MENG S, et al. Soil properties and microbial community in the rhizosphere of *Populus alba* var. *pyramidalis* along a chronosequence [J/OL]. *Microbiological research*, 2021, 250: 126812 [2022-11-03]. <https://doi.org/10.1016/j.micres.2021.126812>.
- [3] FAN J L, ZHENG J, WU L F, et al. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models[J/OL]. *Agricultural water management*, 2021, 245: 106547 [2022-11-03]. <https://doi.org/10.1016/j.agwat.2020.106547>.
- [4] LIU M H, SHI H B, PAREDES P, et al. Estimating and partitioning maize evapotranspiration as affected by salinity using weighing lysimeters and the SIMDualKc model[J/OL]. *Agricultural water management*, 2022, 261: 107362 [2022-11-03]. <https://doi.org/10.1016/j.agwat.2021.107362>.
- [5] CHEN H, JIANG A Z, HUANG J J, et al. An enhanced shuttleworth-wallace model for simulation of evapotranspiration and its components[J/OL]. *Agricultural and forest meteorology*, 2022, 313: 108769 [2022-11-03]. <https://doi.org/10.1016/j.agrformet.2021.108769>.
- [6] MA J, GUO J B, WANG Y H, et al. Variations in stem radii of *Larix principis-rupprechtii* to environmental factors at two slope locations in the Liupan Mountains, Northwest China[J]. *Journal of forestry research*, 2021, 32(2): 513-527.
- [7] TU J, WEI X H, HUANG B B, et al. Improvement of sap flow estimation by including phenological index and time-lag effect in back-propagation neural network models [J/OL]. *Agricultural and forest meteorology*, 2019, 276-277: 107608 [2022-11-03]. <https://doi.org/10.1016/j.agrformet.2019.06.007>.
- [8] 范俊楠, 张钰, 贺小敏, 等. 基于BP神经网络的重点行业企业周边土壤重金属污染预测及评价[J]. *华中农业大学学报*, 2019, 38(4): 55-62. FAN J N, ZHANG Y, HE X M, et al. BP neural network based prediction and evaluation of heavy metal pollution in soil around the enterprises in key areas of Hubei Province[J]. *Journal of Huazhong Agricultural University*, 2019, 38(4): 55-62 (in Chinese with English abstract).
- [9] KOLANOWSKI K, ŚWIETLIĆKA A, KAPELA R, et al. Multisensor data fusion using Elman neural networks [J]. *Applied mathematics and computation*, 2018, 319: 236-244.
- [10] ELMAN J L. Finding structure in time [J]. *Cognitive science*, 1990, 14(2): 179-211.
- [11] BENEVIDES P, CATALAO J, NICO G. Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors[J/OL]. *Remote sensing*, 2019, 11(8): 966 [2022-11-03]. <https://www.mdpi.com/2072-4292/11/8/966>. DOI: 10.3390/rs11080966.
- [12] WANG W C, CHAU K W, QIU L, et al. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition [J]. *Environmental research*, 2015, 139: 46-54.
- [13] AGHELPOUR P, BAHRAMI-PICHAGHCHI H, KARIMPOUR F. Estimating daily rice crop evapotranspiration in limited climatic data and utilizing the soft computing algorithms MLP, RBF, GRNN, and GMDH[J]. *Complexity*, 2022, 2022: 1-18.
- [14] 韩永贵, 高阳, 韩磊, 等. 基于模糊规则的人工神经网络模拟新疆杨蒸腾耗水[J]. *应用生态学报*, 2020, 31(5): 1525-1534. HAN Y G, GAO Y, HAN L, et al. Modeling water consumption of *Populus bolleana* by artificial neural network based on fuzzy rules[J]. *Chinese journal of applied ecology*, 2020, 31(5): 1525-1534 (in Chinese with English abstract).
- [15] DAI Y, ZHENG X J, TANG L S, et al. Stable oxygen isotopes reveal distinct water use patterns of two *Haloxylon* species in the Gurbantonggut Desert[J]. *Plant and soil*, 2015, 389(1): 73-87.
- [16] 李传金, 胡顺军, 郑博文. 古尔班通古特沙漠南缘梭梭(*Haloxylon ammodendron*)群落能量平衡及蒸散特征[J]. *生态学报*, 2021, 41(1): 92-100. LI C J, HU S J, ZHENG B W. Energy balance and evapotranspiration characteristics of *Haloxylon ammodendron* community in the southern margin of the Gurbantonggut Desert[J]. *Acta ecologica sinica*, 2021, 41(1): 92-100 (in Chinese with English abstract).
- [17] GARSON G D. Interpreting neural-network connections[J]. *AI expert*, 1991, 6(4): 46-51.
- [18] 杜斌, 胡笑涛, 王文娥, 等. 交替沟灌玉米灌浆期茎流影响因子敏感性分析与模型适用性研究[J]. *中国农业科学*, 2018, 51(2): 233-245. DU B, HU X T, WANG W E, et al. Stem flow influencing factors sensitivity analysis and stem flow model applicability in filling stage of alternate furrow irrigated maize[J]. *Scientia agricultura sinica*, 2018, 51(2): 233-245 (in Chinese with English abstract).
- [19] 贾国栋, 陈立欣, 李瀚之, 等. 北方土石山区典型树种耗水特征

- 及环境影响因子[J].生态学报,2018,38(10):3441-3452.JIA G D, CHEN L X, LI H Z, et al. The effect of environmental factors on plant water consumption characteristics in a northern rocky mountainous area[J]. Acta ecologica sinica, 2018, 38(10): 3441-3452 (in Chinese with English abstract).
- [20] KARSOLIYA S. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture[J]. International journal of engineering trends and technology, 2012, 6(3): 714-717.
- [21] CHEN X, ZHAO P, HU Y T, et al. Canopy transpiration and its cooling effect of three urban tree species in a subtropical city-Guangzhou, China [J/OL]. Urban forestry & urban greening, 2019, 43: 126368 [2022-11-03]. <https://doi.org/10.1016/j.ufug.2019.126368>.

## Estimating daily transpiration of *Populus alba* var. *pyramidalis* Bunge based on MLR and artificial neural network

XUE Bing<sup>1</sup>, WANG Qijie<sup>1</sup>, MA Bin<sup>2</sup>, LIANG Xue<sup>1</sup>, HOU Zhen'an<sup>1</sup>, JIANG Yan<sup>1</sup>

1. College of Agronomy, Shihezi University, Shihezi 832003, China;

2. Seed Station of Tekesi County Bureau of Agriculture and Rural Affairs, Yili 835500, China

**Abstract** A traditional multiple linear regression model (MLR) and an artificial neural network model (Back propagation (BP) and Elman) were constructed using 8 combinations of 7 meteorological factors including sunshine duration, wind speed, relative humidity, saturated vapor pressure difference, minimum temperature, maximum temperature, and average daily temperature as model inputs to solve the problems of low accuracy in estimating the daily transpiration ( $T_r$ ) of *Populus alba* var. *pyramidalis* Bunge, the main tree species of shelterbelt forests in arid areas, and the poor generalization ability of estimation models. The  $T_r$  value of *Populus alba* var. *pyramidalis* Bunge in 2020 growth season was estimated. The results of estimating three different input combinations of models were compared and evaluated. At the same time, the relative contribution rates of various meteorological factors to the estimated  $T_r$  values were quantified based on the Garson algorithm. The results showed that the accuracy of BP and Elman models in estimating  $T_r$  exceeded 73.66%. Under different input combinations, the estimation accuracy of the artificial neural network model had increased by 8.45%-31.33% compared to the MLR model. Among them, the Elman model with a topological structure of 6-4-4-1 had the highest accuracy of estimation. The relative contribution rate of saturated vapor pressure difference to  $T_r$  estimation was the largest, followed by relative humidity. The relative contribution rate of different temperature variables to the estimated  $T_r$  values was in the increasing order of average daily temperature > minimum temperature > maximum temperature. It is indicated that the neural network model for estimating daily transpiration of *Populus alba* var. *pyramidalis* Bunge can improve the accuracy of estimating the transpiration of shelter forests in arid areas. It will provide scientific guidance for the sustainable development of shelter forests and the precise regulation of agricultural water resources.

**Keywords** estimation of transpiration; multiple linear regression; artificial neural network; *Populus alba* var. *pyramidalis* Bunge; shelter forest; precise regulation of water resources

(责任编辑:赵琳琳)