

马 钊,任传栋,刘 静,等. 基于不同 LSTM 模型和 Hargreaves 模型估算鲁中地区参考作物蒸散量[J].江苏农业学报,2022,38(6):1559-1568.

doi: 10.3969/j.issn.1000-4440.2022.06.014

基于不同 LSTM 模型和 Hargreaves 模型估算鲁中地区参考作物蒸散量

马 钊¹, 任传栋¹, 刘 静¹, 王志真²

(1.山东省水利勘测设计院有限公司,山东 济南 250013; 2.山东省农业交流合作中心,山东 济南 250013)

摘要: 为找出在仅使用温度这一气象参数条件下适用于区域参考作物蒸散量(ET_0)估算的简化模型,本研究以鲁中地区为研究区域,基于 6 个气象站点 1961–2019 年的逐日气象数据,以长短期记忆神经网络(LSTM)模型和 Hargreaves(HS)模型为基础,利用粒子群优化 LSTM 模型(PSO-LSTM)、遗传算法优化 LSTM 模型(GA-LSTM)、贝叶斯理论优化 LSTM 模型(BA-LSTM)和 5 种 HS 改进模型估算 ET_0 ,并将估算结果与 Penman-Monteith(PM)模型的 ET_0 进行对比。结果表明,相同参数输入条件下,LSTM 模型精度普遍优于 HS 模型,4 种 LSTM 模型具有较强的适用性,其中 BA-LSTM 模型对 ET_0 日值的估算效果最优,其均方根差(RMSE)、平均绝对误差(MAE)、决定系数(R^2)、效率系数(E_{ns})的中位线分别为 0.378 mm/d、0.276 mm/d、0.904 和 0.902,其综合性指标指数(GPI)的中位线为 1.837。同时,BA-LSTM 模型在全区的相对误差(RE)仅为 0.01%~1.75%。因此,在仅有温度这一气象参数时,推荐使用 BA-LSTM 模型估算鲁中地区 ET_0 。

关键词: 鲁中地区; 参考作物蒸散量; 温度; 长短期记忆神经网络(LSTM)模型; 贝叶斯理论; Hargreaves 模型

中图分类号: S274 **文献标识码:** A **文章编号:** 1000-4440(2022)06-1559-10

Estimation of reference crop evapotranspiration in central Shandong by different LSTM models and Hargreaves models

MA Zhao¹, REN Chuan-dong¹, LIU Jing¹, WANG Zhi-zhen²

(1.Shandong Survey and Design Institute of Water Conservancy Co., Ltd., Jinan 250013, China; 2.Shandong Agricultural Exchange and Cooperation Center, Jinan 250013, China)

Abstract: To find out a simplified model suitable for regional reference crop evapotranspiration (ET_0) estimation using only temperature data, we took the central region of Shandong as the research area in this study, and used the daily meteorological data of six meteorological stations from 1961 to 2019. Based on long short-term memory neural network (LSTM) model and Hargreaves (HS) model, three optimization models including particle swarm optimization LSTM (PSO-LSTM) model, genetic algorithm optimization LSTM (GA-LSTM) model and Bayesian theory optimization LSTM (BA-LSTM) model and five improved HS models were used to estimate ET_0 . The calculation results were compared with the Penman-Monteith (PM) model. The results showed that under the same parameter input conditions, the accuracy of the

LSTM model was generally better than that of the HS model, and the four LSTM models had strong applicability. The BA-LSTM model had the best estimation effect on the daily value of ET_0 . The medians of root mean square error (RMSE), mean absolute error (MAE), determination coefficient (R^2) and efficiency coefficient (E_{ns}) were 0.378

收稿日期:2022-03-01

基金项目:河北省水利科研和推广项目(202065)

作者简介:马 钊(1980–),男,山东菏泽人,学士,高级工程师,主要从事水利工程规划设计研究。(E-mail)mazhao6842@126.com

通讯作者:任传栋,(E-mail)253510814@qq.com

mm/d, 0.276 mm/d, 0.904 and 0.902. The median of global performance index (GPI) was 1.837. The relative error (RE) of the BA-LSTM model in the whole area was only 0.01%–1.75%. Therefore, it is recommended to use BA-LSTM model to estimate ET_0 in central Shandong when only temperature data are available.

Key words: central region of Shandong; reference crop evapotranspiration; temperature; long short-term memory neural network (LSTM) model; Bayesian theory; Hargreaves model

随着全球变暖现象的日益加剧,区域干旱发生频率逐渐提高,这严重影响居民生活及人身安全^[1-2]。同时,水资源供需矛盾的逐渐加剧,给区域农业发展带来了严峻挑战,严重影响了区域粮食产量及农业产值^[3-4]。参考作物蒸散量(ET_0)是影响区域水资源平衡和能量传递的重要参数之一,其值的准确估算对区域经济发展有着十分重要的意义^[5]。Penman-Monteith (PM) 模型是计算 ET_0 的标准模型,但利用该模型计算 ET_0 时需用到温度、日照时数、相对湿度、风速等多个气象数据,使得该模型在气象资料缺乏地区的应用受限^[6-7]。因此,在保证 ET_0 估算精度的前提下,找寻区域 ET_0 简化计算模型成为国内外的研究热点。

Hargreaves 等^[8]基于温度资料及大气顶层辐射构建了 Hargreaves (HS) 模型,该模型在仅需要温度资料的条件下就能以较高精度估算区域 ET_0 ,但该模型中的经验参数需针对不同区域进行属地化处理,因此许多学者对 HS 模型进行了参数校正。胡庆芳等^[9]基于全局优化的洗牌算法对中国不同区域 HS 模型参数进行了校正,得到了全国不同分区的 HS 模型参数参考值。李晨等^[10]基于贝叶斯原理在川中丘陵区对 HS 模型进行了参数校正,指出改进后的 HS 模型精度明显提高。夏兴生等^[11]基于回归修正原理在中国农业区对 HS 模型进行了修正,同样取得了较高的精度。

近年来,机器学习算法凭借计算的简便性及高精度被广泛应用于区域 ET_0 估算中。Tabari 等^[12]基于支持向量机模型估算了伊朗半干旱区的 ET_0 ,指出模型精度高于 HS、Priestley-Taylor 和 Blaney-Cridle 等经验模型。武剑飞等^[13]在气象资料受限条件下基于 NARX 模型对攀枝花站点 ET_0 进行了估算,指出了该模型精度普遍优于 HS 模型。张薇等^[14]基于梯度提升决策树、随机森林、回归树 3 种机器学习模型估算江苏省 ET_0 ,并将计算结果与 3 种经验模型对比,指出梯度提升决策树精度最高,可作为江苏省简化 ET_0 计算模型使用。不同研究结果均表

明,相同气象参数输入条件下,机器学习模型的精度普遍优于经验模型。

长短期记忆神经网络 (LSTM) 模型将时间概念引入模型结构中,已被广泛应用于数据预测和模拟中^[15-16]。但传统的 LSTM 模型精度易受到参数取值的影响,因此,需采用智能算法对其进行优化。本研究拟基于贝叶斯理论 (BA)、粒子群算法 (PSO) 和遗传算法 (GA) 3 种算法优化 LSTM 模型,选取鲁中地区 6 个气象站点,在仅输入温度数据的条件下,估算区域 ET_0 ,将 3 种优化模型的计算结果与传统 LSTM 模型、传统 HS 模型和 5 种 HS 优化模型进行对比,在气象数据受限条件下,得出区域 ET_0 简化估算最优模型,以期气象数据缺乏地区灌溉制度的制订提供科学依据。

1 材料与方法

1.1 研究区域概况及数据来源

鲁中地区地处山东省中部,属于暖温带大陆性季风气候区,区域四季分明,主要地形包括山区及平原,是山东省重要的粮食生产基地^[17]。为得出鲁中地区 ET_0 估算的最优简化模型,本研究选择鲁中地区济南、泰山、潍坊、淄博、淄川、沂源 6 个气象站点 1961–2019 年的逐日气象数据,数据包括日最高气温 (T_{\max})、日最低气温 (T_{\min})、日照时数 (n)、相对湿度 (RH) 和离地面 10 m 高处风速 (U_{10}),气象数据均来自国家气象中心,数据质量控制良好。研究区域地形及站点分布情况见图 1。

1.2 研究方法

1.2.1 Penman-Monteith 模型 1998 年 FAO-56 分册中规定 Penman-Monteith (PM) 模型为计算 ET_0 的标准模型^[18],具体公式为:

$$ET_0 = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T + 273} U_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 U_2)} \quad (1)$$

式中, R_n 为作物表面的净辐射 [$\text{MJ}/(\text{m}^2 \cdot \text{d})$]; T 为离地面 2 m 高处的日均气温 ($^{\circ}\text{C}$); G 为土壤热

量通量密度 [$\text{MJ}/(\text{m}^2 \cdot \text{d})$]; U_2 为离地面 2 m 高处风速 (m/s); e_s 为饱和水汽压差 (kPa); e_a 为实际水汽压差 (kPa); Δ 为蒸汽压曲线的斜率 ($\text{kPa}/^\circ\text{C}$); γ 为干湿计常数 ($\text{kPa}/^\circ\text{C}$)。

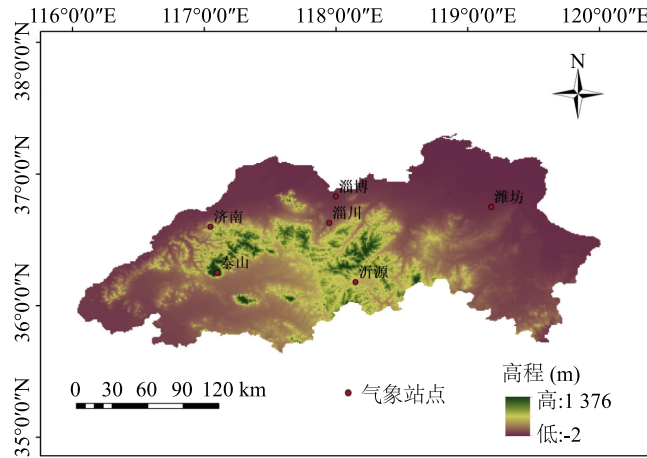


图 1 研究区域概况及站点分布

Fig.1 Research area overview and site distribution

1.2.2 长短期记忆神经网络模型 长短期记忆神经网络 (LSTM) 模型引入时间序列概念,使模型识别训练数据集的时间规律,保证了该模型的精度。模型由输入门、遗忘门、更新门和输出门 4 个部分组成,具体步骤可见文献[19]。

1.2.3 贝叶斯优化 LSTM 模型 贝叶斯优化基于高斯回归原理,对模型进行参数寻优。假设待优化的参数组合为 $X = \{x_1, x_2, \dots, x_n\}$,模型允许的计算输出结果误差函数为 $f(x)$,在预测过程中判定输出参数能否符合条件,其误差函数符合二项分布:

$$f(x) = \text{GP}[\mu, k(x, x)] \quad (2)$$

式中,GP 为高斯分布; μ 为序列均值; $k(x, x)$ 为序列协方差。

通过贝叶斯优化的第 $t-1$ 组数据为 $[x_{t-1}, f(x_{t-1})]$,则第 t 组数据的预测值 $f(x_t)$ 为符合高斯分布的一个样本,公式为:

$$\begin{bmatrix} f_{1:t-1} \\ f_t \end{bmatrix} = \text{GP} \left\{ \mu, \begin{bmatrix} K, k^t \\ k, k(x_t, x_t) \end{bmatrix} \right\} \quad (3)$$

$$K = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_{t-1}) \\ \vdots & & \vdots \\ k(x_{t-1}, x_1) & \dots & k(x_{t-1}, x_{t-1}) \end{bmatrix} \quad (4)$$

$$k = [(x_t, x_1), (x_t, x_2), \dots, (x_t, x_{t-1})] \quad (5)$$

式中, x_1, x_{t-1}, x_t 为样本参数; μ 为序列均值; f_t 、 $f_{1:t-1}$ 为样本计算值。

联立式(3)至式(5)即可解出第 t 个观测值。在 LSTM 模型训练过程中,通过贝叶斯理论进行迭代计算,得出满足要求的预测值。以原始数据集为起点,通过本研究提出的贝叶斯目标函数及 LSTM 模型,即可得出优化后的 LSTM 模型 (BA-LSTM)。

1.2.4 粒子群算法优化 LSTM 模型 粒子群算法 (PSO) 基于群鸟喂养的原理,将每个训练数据集视为一个粒子,全部数据集即为粒子群,应用每个粒子本身的运动属性完成算法参数寻优过程^[20]。在参数迭代过程中,算法通过计算数据个体的极值从而更新粒子运动属性,直到参数满足误差要求为止。将 PSO 算法用于优化 LSTM 模型,构建 PSO-LSTM 模型,可快速选取最优参数,提高模型的精度。具体步骤可参照文献[21]。

1.2.5 遗传算法优化 LSTM 模型 遗传算法 (GA) 通过交叉、变异处理计算训练数据的个体适应度,通过比较个体适应度的值能否满足要求,判定是否输出最终结果。该算法基于生物界适者生存的原理,保留最优个体,从而提高模型参数的迭代速率。构建的 GA-LSTM 模型可在传统 LSTM 模型的基础上提高模型的运行速率和精度,模型具体步骤可见参考文献[22]。

1.2.6 Hargreaves 模型 Hargreaves (HS) 模型仅基于温度资料便能以较高精度估算区域 ET_0 ^[23],具体

公式为:

$$ET_0 = 0.0023 [(T_{\max} - T_{\min})^{0.5000}] (T_{\text{mean}} + 17.8000) \frac{R_a}{\lambda} \quad (6)$$

式中, ET_0 表示参考作物蒸散量 (mm/d); T_{\max} 、 T_{\min} 、 T_{mean} 分别表示日最高温度 (°C)、日最低温度 (°C)、日平均温度 (°C); R_a 表示大气顶层辐射 [MJ/(m²·d)]; λ 表示汽化潜热常数。

1.2.7 Droogres 改进的 Hargreaves 模型 Droogres 等^[24] 基于气象数据观测的随机性, 提出了 2 种 Hargreaves 改进模型, 分别为 HSM1、HSM2 模型, 具体公式分别为:

$$ET_0 = 0.4080 \times 0.0030 \times [(T_{\max} - T_{\min})^{0.4000}] (T_{\text{mean}} + 20.0000) R_a \quad (7)$$

$$ET_0 = 0.4080 \times 0.0025 \times [(T_{\max} - T_{\min})^{0.5000}] (T_{\text{mean}} + 16.8000) R_a \quad (8)$$

式中, ET_0 表示参考作物蒸散量 (mm/d); T_{\max} 、 T_{\min} 、 T_{mean} 分别表示日最高温度 (°C)、日最低温度 (°C)、日平均温度 (°C); R_a 表示大气顶层辐射 [MJ/(m²·d)]。

1.2.8 胡庆芳等改进的 Hargreaves 模型 胡庆芳等^[25] 基于洗牌算法对全国不同分区 HS 模型参数进行了校正, 改进的模型为 HSM3 模型, 基本公式为:

$$ET_0 = 0.0010 \times R_a [(T_{\max} - T_{\min})^{0.6600}] (T_{\text{mean}} + 34.5000) \quad (9)$$

式中, ET_0 表示参考作物蒸散量 (mm/d); T_{\max} 、 T_{\min} 、 T_{mean} 分别表示日最高温度 (°C)、日最低温度 (°C)、日平均温度 (°C); R_a 表示大气顶层辐射 [MJ/(m²·d)]。

1.2.9 夏兴生等改进的 Hargreaves 模型 夏兴生等^[11] 对中国农业区不同分区 HS 模型进行了回归修正, 改进的模型为 HSM4 模型, 基本公式为:

$$ET_0 = a + bET_{0-\text{HS}} \quad (10)$$

式中, ET_0 表示参考作物蒸散量 (mm/d); $ET_{0-\text{HS}}$ 表示 HS 模型的参考作物蒸散量计算结果 (mm/d); a 、 b 表示回归经验参数。

1.2.10 基于贝叶斯原理改进的 Hargreaves 模型 李晨等^[10] 基于贝叶斯原理对川中丘陵区 HS 模型参数进行了修正, 改进的模型为 HSM5 模型, 基本公式为:

$$ET_0 = C \times R_a [(T_{\max} - T_{\min})^m] (T_{\text{mean}} + a) \quad (11)$$

式中, ET_0 表示参考作物蒸散量 (mm/d); T_{\max} 、

T_{\min} 、 T_{mean} 分别表示日最高温度 (°C)、日最低温度 (°C)、日平均温度 (°C); R_a 表示大气顶层辐射 [MJ/(m²·d)]; C 、 m 、 a 为修正后的经验参数。本研究同样基于贝叶斯原理修正了鲁中地区 6 个站点 HS 模型的经验参数, 鲁中地区不同站点修正后的参数取值见表 1。

表 1 不同站点 HSM5 模型参数取值

Table 1 Parameter values of HSM5 model at different stations

| 站点 | 参数 C | 参数 m | 参数 a |
|----|---------|--------|--------|
| 济南 | 0.00239 | 0.129 | 17.794 |
| 泰山 | 0.00230 | 0.259 | 17.792 |
| 潍坊 | 0.00243 | 0.074 | 17.790 |
| 淄博 | 0.00229 | 0.129 | 17.790 |
| 淄川 | 0.00237 | 0.047 | 17.793 |
| 沂源 | 0.00238 | 0.049 | 17.789 |

1.2.11 模型训练与验证 仅使用温度数据, 以最高温度 (T_{\max}) 和最低温度 (T_{\min}) 为模型输入数据, 以 1961–2000 年的数据训练模型, 2001–2019 年的数据验证模型, 不同模型及算法的参数取值见表 2。以均方根差 (RMSE)、决定系数 (R^2)、平均绝对误差 (MAE)、效率系数 (E_{ns}) 和综合性指标指数 (GPI) 评价不同模型精度, 计算公式分别为:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - X_i)^2} \times 100\% \quad (12)$$

$$R^2 = \frac{[\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^m (X_i - \bar{X})^2 \sum_{i=1}^m (Y_i - \bar{Y})^2} \quad (13)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i| \quad (14)$$

$$E_{\text{ns}} = 1 - \frac{\sum_{i=1}^m (Y_i - X_i)^2}{\sum_{i=1}^m (X_i - \bar{X})^2} \quad (15)$$

$$GPI_i = \sum_{j=1}^4 [\alpha_j (g_j - y_{ij})] \quad (16)$$

式中, m 为逐日数据个数; X_i 、 Y_i 分别为模型模拟值和 PM 模型标准值; \bar{X} 和 \bar{Y} 分别为 X_i 和 Y_i 的平均值; α_j 为常数, 计算 MAE 和 RMSE 时取 1, 计算 E_{ns} 和 R^2 时取 -1; g_j 为不同指标的中位数; y_{ij} 为不同指标的计算值。

2 结果与分析

2.1 不同模型估算的 ET_0 日值拟合效果分析

将鲁中地区不同站点不同模型估算的 ET_0 日值与 PM 模型 ET_0 日值进行对比,表 3 显示,不同模型精度不同,在相同气象参数输入的条件下,LSTM 模型精度普遍优于 HS 模型及 HS 改进模型。BA-LSTM 模型估算的 ET_0 日值在不同站点与 PM 模型的 ET_0 日值拟合效果最好,其在不同站点的拟合方

程斜率最接近标准值 1,决定系数 R^2 均在 0.850 以上。PSO-LSTM 模型和 GA-LSTM 模型估算的 ET_0 日值与 PM 模型的 ET_0 日值拟合效果次之,在不同站点的拟合方程斜率分别为 0.775~0.979 和 0.746~0.978, R^2 分别为 0.795~0.919 和 0.754~0.915。在经验模型中,HSM5 模型估算的 ET_0 日值拟合效果最优,其在不同站点的拟合方程斜率为 0.710~0.901,决定系数 R^2 在 6 种经验模型中较高。

表 2 不同算法参数取值情况

Table 2 Parameter values of different algorithms

| 算法 | 参数取值 |
|-------------|--|
| 粒子群算法 | 微粒种群数目 $N=50$,加速度因子 $c_1=c_2=1.5$,最大速度 $V_{\max}=1.0$,惯性权重 $\omega=0.5$ |
| 遗传算法 | 隐含层节点数 $n=10$,迭代次数 $N=50$,种群规模 $s=10$,交叉概率 $P_c=0.4$,变异概率 $P_m=0.2$ |
| 贝叶斯理论算法 | 迭代次数取 29,初始学习率 0.001,随机梯度下动量 0.801,正则化系数 0.001 488 |
| 长短期记忆神经网络模型 | 最大训练周期数取 1,分块尺寸为 50,初始学习速率为 0.001 |

表 3 不同模型估算的不同站点 ET_0 日值与 PM 模型 ET_0 日值拟合效果对比

Table 3 Comparison of fitting effects between daily ET_0 value estimated by different models at different sites and daily ET_0 value estimated by PM model

| 站点 | PSO-LSTM | | BA-LSTM | | GA-LSTM | | LSTM | | HS | | HSM1 | | HSM2 | | HSM3 | | HSM4 | | HSM5 | |
|----|----------|-------|---------|-------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 | 斜率 | R^2 |
| 济南 | 0.908 | 0.902 | 0.913 | 0.909 | 0.884 | 0.877 | 0.865 | 0.875 | 1.649 | 0.676 | 0.814 | 0.682 | 0.803 | 0.686 | 1.517 | 0.681 | 1.602 | 0.680 | 0.831 | 0.699 |
| 泰山 | 0.775 | 0.795 | 0.812 | 0.851 | 0.746 | 0.754 | 0.720 | 0.749 | 1.471 | 0.470 | 0.657 | 0.450 | 0.638 | 0.468 | 1.350 | 0.512 | 1.390 | 0.470 | 0.710 | 0.604 |
| 潍坊 | 0.943 | 0.919 | 0.948 | 0.921 | 0.921 | 0.915 | 0.817 | 0.917 | 1.563 | 0.692 | 1.197 | 0.781 | 1.216 | 0.791 | 1.385 | 0.713 | 1.478 | 0.712 | 0.806 | 0.787 |
| 沂源 | 0.955 | 0.910 | 0.957 | 0.916 | 0.935 | 0.905 | 0.891 | 0.894 | 1.740 | 0.627 | 1.174 | 0.720 | 1.193 | 0.726 | 1.345 | 0.703 | 1.339 | 0.707 | 0.890 | 0.734 |
| 淄博 | 0.877 | 0.891 | 0.888 | 0.913 | 0.849 | 0.855 | 0.836 | 0.851 | 1.716 | 0.663 | 0.802 | 0.675 | 0.797 | 0.672 | 1.598 | 0.670 | 1.545 | 0.663 | 0.834 | 0.742 |
| 淄川 | 0.979 | 0.906 | 0.982 | 0.917 | 0.978 | 0.883 | 0.962 | 0.879 | 1.886 | 0.620 | 1.229 | 0.711 | 1.257 | 0.719 | 1.488 | 0.723 | 1.453 | 0.720 | 0.901 | 0.733 |

LSTM:传统长短期记忆神经网络模型;PSO-LSTM:粒子群算法优化的 LSTM 模型;BA-LSTM:贝叶斯优化的 LSTM 模型;GA-LSTM:遗传算法优化的 LSTM 模型;HS:Hargreaves 模型;HSM1:Droogres 等^[24]优化的 HS 模型;HSM2:Droogres 等^[24]优化的 HS 模型;HSM3:胡庆芳等^[25]优化的 HS 模型;HSM4:夏兴生等^[11]优化的 HS 模型;HSM5:本研究基于贝叶斯原理优化的 HS 模型; R^2 :决定系数。

2.2 不同模型估算的 ET_0 月值拟合效果分析

不同模型不同站点 ET_0 月值与 PM 模型 ET_0 月值的拟合结果(图 2)表明,不同模型估算的 ET_0 月值有所差异,但在不同月份的变化趋势基本一致,均呈抛物线趋势变化。4 种机器学习模型估算的 ET_0 月值与 PM 模型的 ET_0 月值拟合效果优于其余模型,其中 BA-LSTM 模型估算的 ET_0 月值与 PM 模型的拟合效果最优。在经验模型中,HSM5 模型估算的 ET_0 月值与 PM 模型的拟合效果最优,HS 模型计算结果精度较低。HS 模型和 5 种 HS 改进模型基本均高估了 ET_0 月值,HS 模型、HSM3 模型和 HSM4 模型估算的

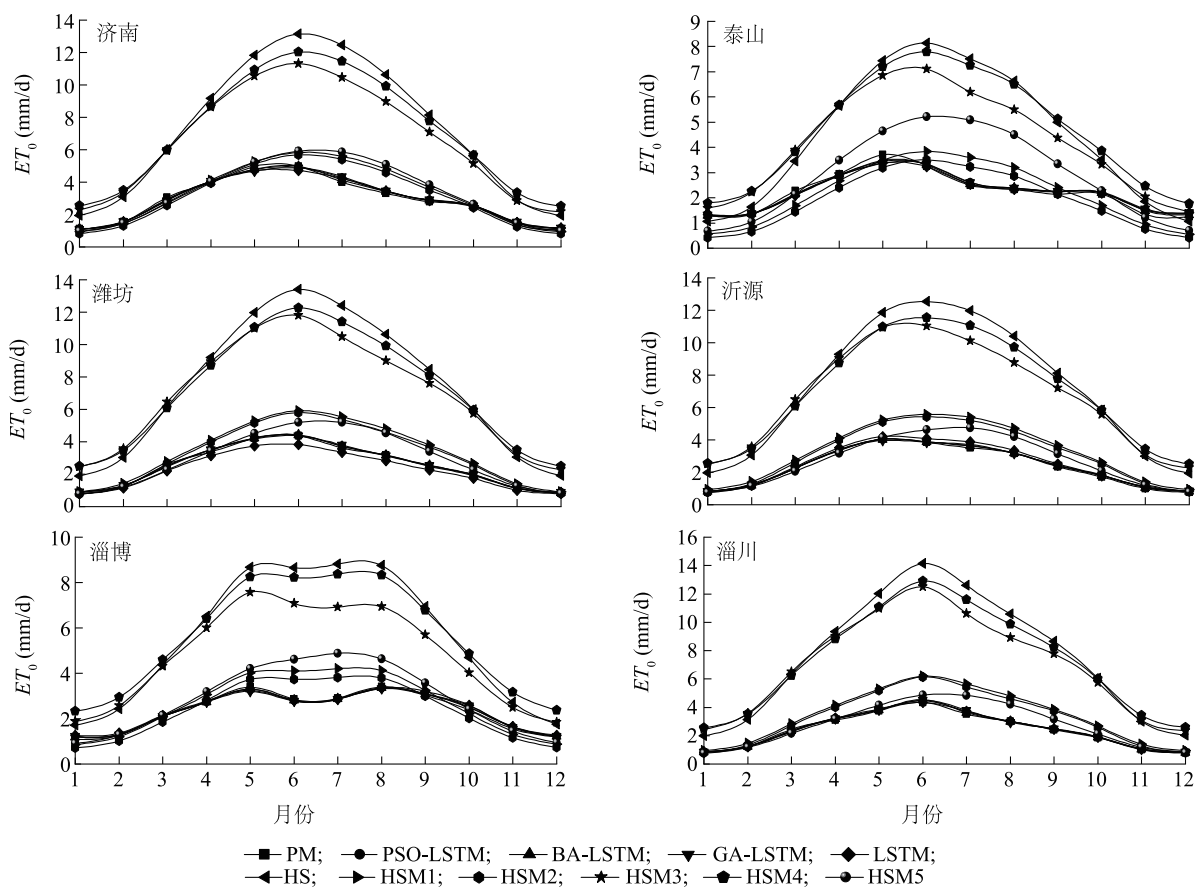
ET_0 月值与 PM 模型的 ET_0 月值拟合效果较差。

2.3 不同模型估算的 ET_0 日值精度指标对比

为进一步比较不同模型的计算精度,本研究比较了不同模型估算的 ET_0 日值,结果(图 3)显示,4 种 LSTM 模型估算的 ET_0 日值比 6 种 HS 模型估算结果精度更高,综合考虑,在所有模型中 BA-LSTM 模型估算的 ET_0 日值精度最高,在 6 种经验模型中 HSM5 模型估算的 ET_0 日值精度最高。在 ET_0 日值计算结果对比中,PM 模型估算的 ET_0 日值中位线为 2.43 mm/d,BA-LSTM 模型估算的 ET_0 日值中位线为 2.45 mm/d,其余 LSTM 模型估算的 ET_0 日值中位线为 2.46~2.48

mm/d; 6 种经验模型中, 优化后的 HS 模型估算的 ET_0 日值精度高于传统 HS 模型, HSM5 模型估算的 ET_0 日值中位线与 PM 模型较接近, 为 2.65 mm/d, 其余 HS 模型估算的 ET_0 日值中位线为 3.15~6.56 mm/d。此外, 从 5 项精度指标综合考虑, BA-LSTM 模型估算的 ET_0 日值精度最高, 其均方根差 (RMSE)、平均绝对误差 (MAE)、决定系数 (R^2)、效率系数 (E_{ns}) 的中位线分别为 0.378 mm/d、0.276 mm/d、0.904 和 0.902, 其 GPI 的中位线为 1.837, 在所有模型中排名第一。PSO-LSTM 模型和 GA-LSTM 模型估算 ET_0 日值的

精度次之, 均优于 LSTM 模型, GPI 中位线分别为 1.724 和 1.566。4 种 LSTM 模型估算 ET_0 日值的精度高于 6 种 HS 模型, 在经验模型中, HSM5 模型估算的 ET_0 日值精度最高, 其 RMSE、MAE、 R^2 、 E_{ns} 的中位线分别为 1.122 mm/d、0.838 mm/d、0.716 和 0.609, GPI 中位线为 0.166, 在所有模型中排名第 5 位, 其次为 HSM1、HSM2、HSM3、HSM4 模型, 传统 HS 模型估算 ET_0 日值的精度最低。因此, 在仅输入温度这一气象参数条件下, BA-LSTM 模型可用于估算鲁中地区 ET_0 。



ET_0 : 参考作物蒸散量; PM: Penman-Monteith 模型。LSTM、PSO-LSTM、BA-LSTM、GA-LSTM、HS、HSM1、HSM2、HSM3、HSM4、HSM5 见表 3 注。

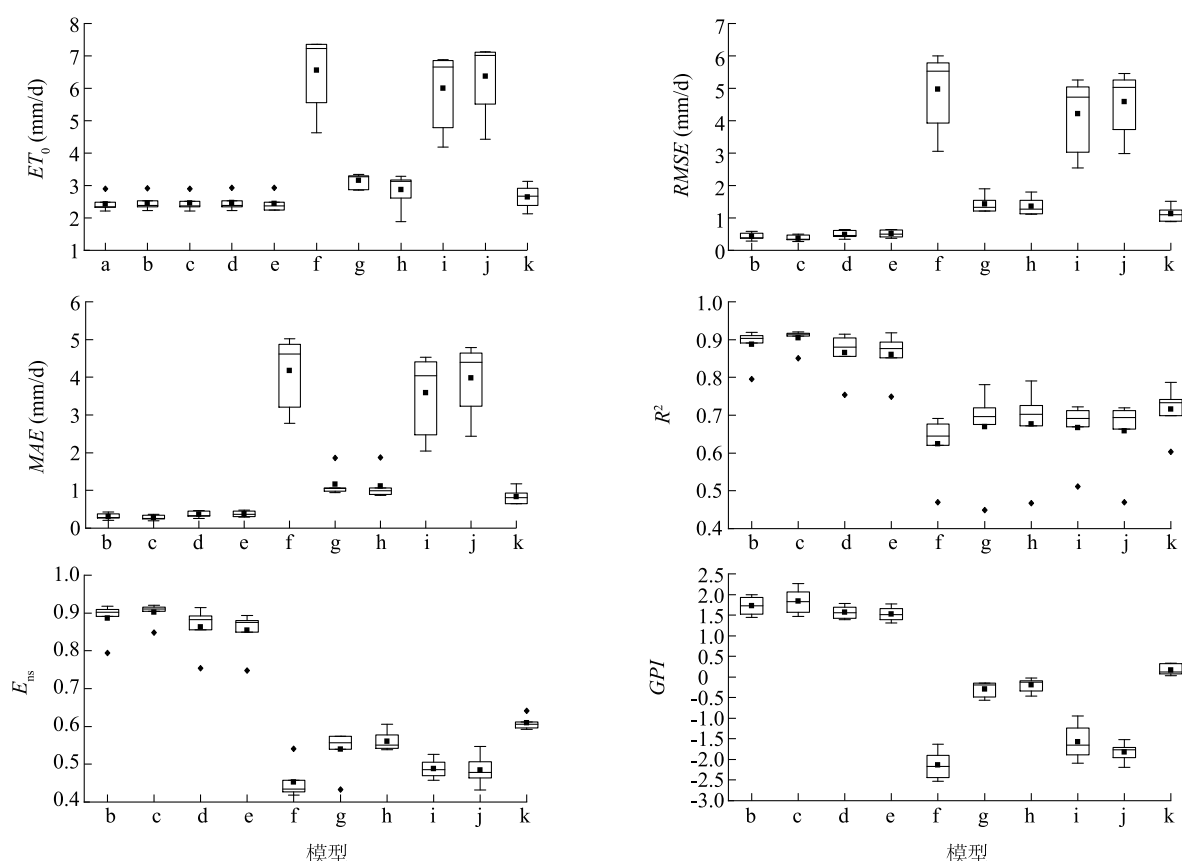
图 2 不同模型估算的 ET_0 月值与 PM 模型 ET_0 月值的拟合对比

Fig.2 Comparison of fitting effects between monthly ET_0 estimated by different models and monthly ET_0 estimated by PM model

2.4 不同模型估算的 ET_0 日值相对误差对比

图 4 显示, 不同模型计算鲁中地区 ET_0 日值的相对误差 (RE) 不同, 但分布趋势基本一致, 均呈现西部相对误差较低, 中部相对误差较高的趋势。其中, BA-LSTM 模型在整个鲁中地区的 RE 均较低, RE 最高值出现在沂源站附近, 仅为 1.75%。

PSO-LSTM 模型和 GA-LSTM 模型精度次之, RE 最高值分别为 2.01% 和 2.42%。在经验模型中, HSM5 模型的 RE 较低, 为 2.08%~16.96%, 其余改进的 HS 模型的精度较低, 但均优于传统 HS 模型。HSM5 模型可在保证计算简便的前提下, 明显提高模型计算精度, 机器学习模型的精度普遍优



a; PM; b; PSO-LSTM; c; BA-LSTM; d; GA-LSTM; e; LSTM; f; HS; g; HSM1; h; HSM2; i; HSM3; j; HSM4; k; HSM5. PM: Penman-Monteith 模型。LSTM、PSO-LSTM、BA-LSTM、GA-LSTM、HS、HSM1、HSM2、HSM3、HSM4、HSM5 见表 3 注。 ET_0 : 参考作物蒸散量; $RMSE$: 均方根差; MAE : 平均绝对误差; R^2 : 决定系数; E_{ns} : 效率系数; GPI : 综合性指标指数。

图 3 不同模型 ET_0 日值精度箱线图

Fig.3 Boxplots of daily ET_0 accuracy of different models

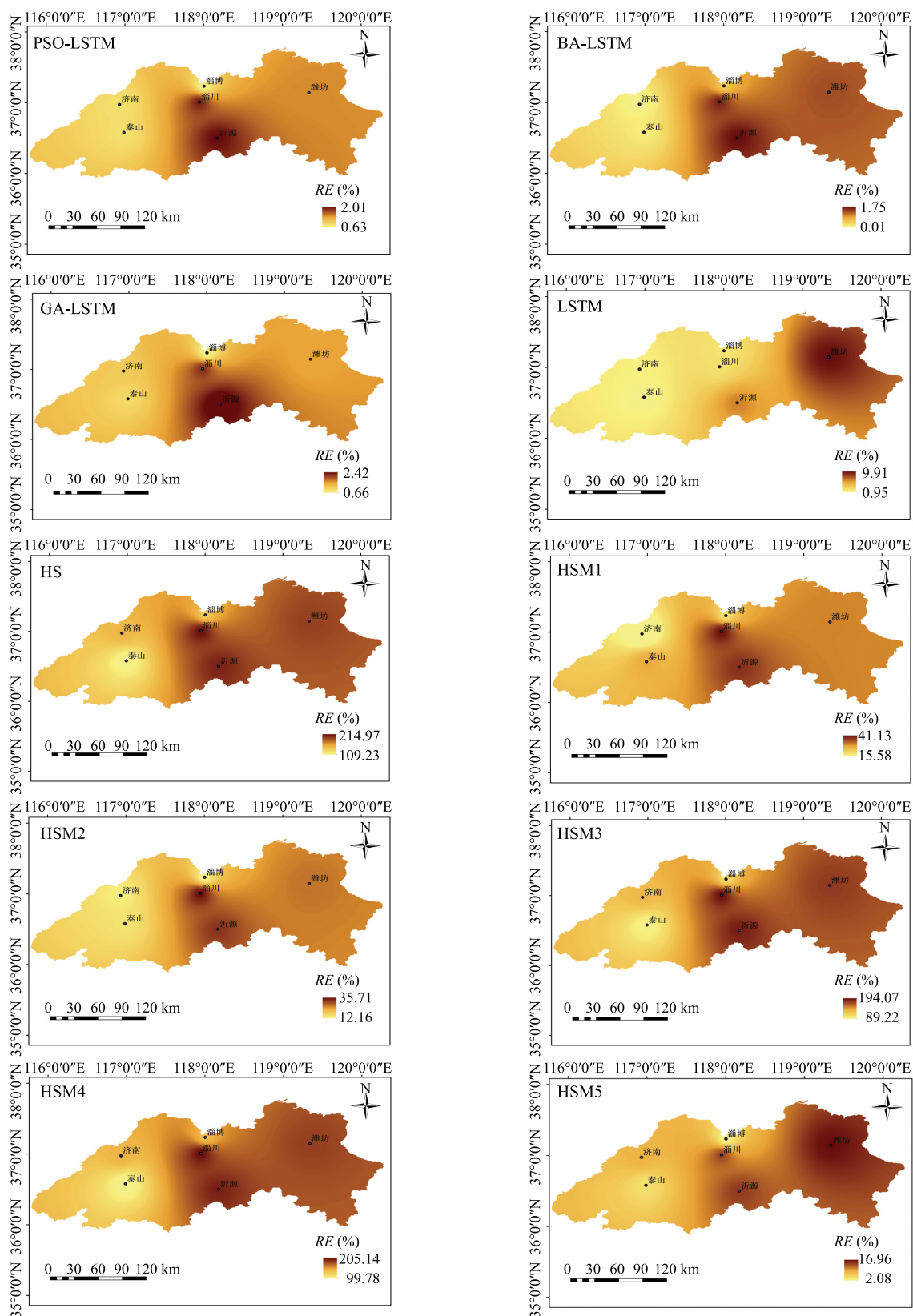
于 HS 模型, BA-LSTM 模型在整个区域均可保证较高的精度。

3 讨论

温度资料是气象数据中最容易获得的, 并且其观测精度较高, HS 模型仅需温度资料即可估算区域 ET_0 。本研究发现, 未优化的 HS 模型在鲁中地区的计算精度较低, 这说明 HS 模型在应用时需要参数进行属地化处理。HSM5 模型基于贝叶斯原理, 可将新计算得出的数据作为后验分布, 将后验分布作为新一轮计算的先验分布, 以此保证模型计算的精确性。冯禹等^[26]在四川盆地比较了不同 HS 改进模型的计算精度, 指出基于贝叶斯原理优化的 HS 模型计算精度最高; 赵凡萱等^[27]在宁夏地区比较了贝叶斯优化的 HS 模型与其他修正模型的计算精度, 得出的结论与本研究结果基本一致。

LSTM 模型作为深度学习模型的一种^[28], 基于传统机器学习模型, 在模型训练时引入时间序列, 可以更好地反映训练数据在时间尺度上的变化规律, 摆脱传统机器学习模型参数随机性选取的缺点, 降低了模型训练的随机性。从本研究的箱线图中可以看出, 4 种 LSTM 模型精度指标的标准差较小, 表明 LSTM 模型无需进行参数属地化处理即可应用于区域 ET_0 的估算。李莉等^[29]在研究番茄蒸散量估算时比较了 LSTM 模型与 Elman 模型、RNN 模型的精度, 指出 LSTM 模型具有较高的预测精度和泛化性能。谢家兴等^[30]比较了 LSTM 模型、ELM 模型、GRNN 模型的精度, 也得出了相同的结论。

本研究结果表明, 经算法优化的 LSTM 模型可进一步提高精度, 其中, BA-LSTM 模型精度最优。PSO 算法和 GA 算法已被证明可用于区域 ET_0 估算的机器学习算法优化中, 算法的应用可普遍提高机



RE: 相对误差。LSTM、PSO-LSTM、BA-LSTM、GA-LSTM、HS、HSM1、HSM2、HSM3、HSM4、HSM5 见表 3 注。

图 4 不同模型估算的 ET_0 日值相对误差空间分布对比

Fig.4 Comparison of spatial distribution of relative errors of daily ET_0 estimated by different models

器学习模型精度和迭代速率。贝叶斯理论可在传统模型参数基础上,自动调节模型参数,优化模型结构,提高模型的泛化能力。综上所述,在仅有温度这一气象参数条件下,可推荐使用 BA-LSTM 模型估算鲁中地区 ET_0 。

4 结 论

本研究在仅输入温度数据的条件下,比较了 BA-LSTM、PSO-LSTM、GA-LSTM、LSTM 共 4 种机器学习模型和 6 种 HS 模型的精度,找出适用于鲁中地区 ET_0 的简化估算模型,得出以下结论:

(1) 在 ET_0 日值估算中,4 种 LSTM 模型的估算效果明显优于 6 种 HS 模型,其中,BA-LSTM 模型对 ET_0 日值的估算效果最优,其不同站点的拟合方程斜率为 0.812~0.982,决定系数均在 0.850 以上。

(2) 在 ET_0 月值估算中,不同模型估算结果的变化趋势基本一致,以 BA-LSTM 模型估算的结果与 PM 模型的拟合效果最优。

(3) 从误差指标中可以看出,BA-LSTM 模型估算精度最高 ($RMSE$ 、 MAE 、 R^2 、 E_{ms} 、 GPI 的中位线分别为 0.378 mm/d、0.276 mm/d、0.904、0.902 和 1.837),经验模型中 HSM5 模型估算精度最高 ($RMSE$ 、 MAE 、 R^2 、 E_{ms} 、 GPI 的中位线分别为 1.122 mm/d、0.838 mm/d、0.716、0.609、0.166)。

(4) 从全区来看,不同模型估算结果的 RE 空间分布趋势基本一致,BA-LSTM 模型在全区的估算精度均最高, RE 仅为 0.01%~1.75%。

综上所述,BA-LSTM 模型可作为鲁中地区 ET_0 简化估算的推荐模型。

参考文献:

- [1] 李可利,张 鑫. 基于 ANFIS 的陕西省参考作物蒸散量计算[J]. 自然资源学报,2020,35(6):1472-1483.
- [2] 关 静,梁 川,赵 璐,等. 3 种干旱指数在宁夏中部干旱带的适用性分析[J]. 水土保持研究,2018,25(3):122-127.
- [3] CHEN Z J, ZHU Z C, JIANG H, et al. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods[J]. Journal of Hydrology, 2020, 591:125286.
- [4] 褚荣浩,李 萌,沙修竹,等. 13 种典型参考作物蒸散量估算模型在安徽省的适用性评价[J]. 节水灌溉,2021(9):61-70.
- [5] YANG J Y, LIU Q, MEI X R, et al. Spatiotemporal characteristics of reference evapotranspiration and its sensitivity coefficients to climate factors in Huang-Huai-Hai Plain, China[J]. Journal of Integrative Agriculture, 2013, 12(12): 2280-2291.
- [6] 张戴伟,王春颖,韩宇平,等. 基于温度的 ET_0 模型改进及在海河流域的适用性[J]. 人民黄河,2021,43(3):155-160.
- [7] 李 晨. 川中丘陵区不同 Hargreaves 改进模型适应性评价[J]. 节水灌溉,2021(11):88-96.
- [8] HARGREAVES G H, ALLEN R G. History and evaluation of Hargreaves evapotranspiration equation[J]. Journal of Irrigation and Drainage Engineering, 2003, 129(1): 53-63.
- [9] 胡庆芳,杨大文,王银堂,等. Hargreaves 公式的全局校正及适用性评价[J]. 水科学进展,2011,22(2):160-167.
- [10] 李 晨,崔宁博,魏新平,等. 改进 Hargreaves 模型估算川中丘陵区参考作物蒸散量[J]. 农业工程学报,2015, 31(11):129-135.
- [11] 夏兴生,赵希振,汤 童,等. Hargreaves 公式在中国农业区的年内回归修正[J]. 排灌机械工程学报,2021,39(1):81-88.
- [12] TABARI H, KISI O, EZANI A, et al. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment[J]. Journal of Hydrology, 2012, 444: 78-89.
- [13] 武剑飞,康银红,宋 鑫,等. 基于 NARX 模型的参考作物蒸散发预测[J]. 排灌机械工程学报,2021,39(5):533-540.
- [14] 张 薇,韦 群,吴天傲,等. 基于 GBDT 算法的参考作物蒸散量模型在江苏省的预测[J]. 江苏农业学报,2020,36(5):1169-1180.
- [15] 孙龙清,吴雨寒,孙希蓓,等. 基于 IBAS 和 LSTM 网络的池塘水溶解氧含量预测[J]. 农业机械学报,2021,52(增刊):252-260.
- [16] 邢立文,崔宁博,董 娟. 基于 LSTM 深度学习模型的华北地区参考作物蒸散量预测研究[J]. 水利水电技术,2019,50(4):64-72.
- [17] 环海军,刘 岩,孙燕玲,等. 鲁中地区近 35a 不同历时降水变化特征[J]. 沙漠与绿洲气象,2018,12(5):83-89.
- [18] ALLEN R G, PEREIRA L S, RAES D, et al. Crop evapotranspiration: guidelines for computing crop water requirements[M]. Rome: Food and Agriculture Organization of United Nation, 1998.
- [19] ZHANG J, ZHU Y, ZHANG X, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas[J]. Journal of Hydrology, 2018, 561: 918-929.
- [20] ZHU B, FENG Y, GONG D Z, et al. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data[J]. Computers and Electronics in Agriculture, 2020, 173: 105430.
- [21] 任 妮,鲍 彤,刘 杨,等. 基于粒子群优化算法和长短时记忆神经网络的蟹塘溶解氧预测[J]. 江苏农业学报,2021,37(2):426-434.
- [22] 黄 敏. 基于进化算法的 LSTM 降水量预测方法研究[D]. 阜新:辽宁工程技术大学,2020.
- [23] HARGREAVES G H, ALLEN R G. History and evaluation of Hargreaves evapotranspiration equation[J]. Journal of Irrigation and

- Drainage Engineering, 2003, 129(1): 53-63.
- [24] DROOGRES P, ALLEN R G. Estimating reference evapotranspiration under inaccurate data conditions[J]. Irrigation and Drainage Systems, 2002, 16(1): 33-45.
- [25] 胡庆芳,杨大文,王银堂,等. Hargreaves 公式的全局校正及适用性评价[J]. 水科学进展,2011,22(2):160-167.
- [26] 冯 禹,崔宁博,龚道枝. 机器学习算法和 Hargreaves 模型在四川盆地 ET_0 计算中的比较[J]. 中国农业气象,2016,37(4): 415-421.
- [27] 赵凡萱,唐诗佳,张 卓,等. 修正 Hargreaves-Samani 公式估算宁夏参考作物蒸散量适用性分析[J]. 中国防汛抗旱,2021,31(6):57-63.
- [28] 蒋敏兰,吴沛伦,陈昊然. 基于多层 LSTM 的海兰褐蛋鸡产蛋率回归模型[J]. 江苏农业科学,2020,48(22):248-251.
- [29] 李 莉,李文军,马德新,等. 基于 LSTM 的温室番茄蒸腾量预测模型研究[J]. 农业机械学报,2021,52(10):369-376.
- [30] 谢家兴,高 鹏,孙道宗,等. 基于长短期记忆的柑橘园蒸散量预测模型[J]. 农业机械学报,2020,51(增刊2):351-356.

(责任编辑:王 妮)