

基于机器学习的 Angström- Prescott 公式系数的估算

冯文哲^{1,2}, 和志豪^{1,2}, 陈 上³, 董文彪^{1,2}, 李若彤^{1,2}, 于 强⁴,
冯 浩^{2,4}, 何建强^{1,2}

(1. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西 杨凌 712100; 2. 西北农林科技大学中国旱区节水农业研究院, 陕西 杨凌 712100; 3. 南京信息工程大学江苏省农业气象重点实验室, 南京 210044; 4. 中国科学院水利部水土保持研究所黄土高原土壤侵蚀与旱地农业国家重点实验室, 陕西 杨凌 712100)

摘 要: 地表太阳辐射 (R_s) 数据在水文、农业和生态等领域具有重要的应用价值。由于目前仅有少数国家气象站点具备直接观测条件, 因此 Angström-Prescott (A-P) 公式被广泛应用于逐日 R_s 的估算。尽管使用 A-P 公式需提供的两个经验系数 a 和 b 已经有 FAO (Food and Agriculture Organization) 推荐值 ($a=0.25$; $b=0.5$), 但是越来越多的研究指出这两个参数的本地化有助于提高 R_s 的估算精度。利用 1967–2017 年全国 80 个具有太阳辐射观测数据气象站的逐日地表太阳辐射 (R_s) 及其他常规气象数据, 来获取中国大陆地区 A-P 公式的 a 、 b 系数。首先, 整个中国大陆地区被划分为高原山地气候区 (Mountain Plateau Zone, MPZ)、亚热带季风气候区 (Subtropical Monsoon Zone, SMZ)、温带季风气候区 (Temperate Monsoon Zone, TMZ)、温带大陆性气候区 (Temperate Continental Zone, TCZ) 等 4 个不同气候区。其次, 基于最小二乘法回归得到各气候区不同站点 A-P 公式系数值, 可视为 A-P 公式系数的观测值。然后, 利用 4 种机器学习算法分别估算全国 80 个具有太阳辐射观测数据气象站的 A-P 公式系数, 各算法分别结合不同输入因子组合构建不同的 A-P 公式系数估算模型。最后, 评估机器学习算法估算得到的 A-P 公式系数自身的精度, 及其在 R_s 估算中的精度。研究发现在估算系数 a 时, 机器学习模型中基于五因子输入组合的 SVM 模型的估算精度最高 ($R^2=0.661$, $RMSE=0.022$, $nRMSE=0.120$)。在估算系数 b 时, 机器学习模型中基于四因子输入组合的 ELM 模型的估算精度最高 ($R^2=0.550$, $RMSE=0.031$, $nRMSE=0.055$)。基于所选最优机器学习模型各自估算的 a 和 b 系数值来驱动 A-P 公式进一步估算 R_s , 结果表明机器学习模型在 MPZ、SMZ、TMZ、TCZ 气候区 R_s 估算中的 $nRMSE$ 分别为 0.168、0.225、0.138、0.180。因此, 推荐使用五因子输入组合的 SVM 模型来估算系数 a , 使用四因子输入组合的 ELM 模型来估算系数 b , 可以得到更为准确的中国大陆地区 A-P 公式系数, 从而提高使用 A-P 公式估算 R_s 的精度。研究结果可为实现 A-P 公式系数在中国大陆地区的本地化和提高 R_s 估算精度提供一定的理论依据和技术途径。

关键词: 太阳辐射; Angström-Prescott; a 、 b 系数; 机器学习

中图分类号: S161.1 **文献标识码:** A **DOI:** 10.12396/jsgg.2024016

冯文哲, 和志豪, 陈 上, 等. 基于机器学习的 Angström-Prescott 公式系数的估算[J]. 节水灌溉, 2024(7): 108–118. DOI: 10.12396/jsgg.2024016.

FENG W Z, HE Z H, CHEN S, et al. Estimation of the coefficients of the Angström-Prescott formula based on machine learning methods [J]. Water Saving Irrigation, 2024(7): 108–118. DOI: 10.12396/jsgg.2024016.

收稿日期: 2024-01-10 接受日期: 2024-02-23

作者简介: 冯文哲 (1999-), 男, 硕士研究生, 研究方向为农业生态系统模拟研究。E-mail: wenzhefeng@nwafu.edu.cn.

通讯作者: 何建强 (1977-), 男, 教授, 博士, 研究方向为农业生态系统模拟研究。E-mail: jianqiang_he@nwsuaf.edu.cn.

Estimation of the Coefficients of the Angström–Prescott Formula Based on Machine Learning Methods

FENG Wen-zhe^{1,2}, HE Zhi-hao^{1,2}, CHEN Shang³, DONG Wen-biao^{1,2},
LI Ruo-tong^{1,2}, YU Qiang⁴, FENG Hao^{2,4}, HE Jian-qiang^{1,2}

(1.Key Laboratory for Agricultural Soil and Water Engineering in Arid Area of Ministry of Education, Northwest A&F University, Yangling 712100, Shaanxi Province, China; 2.Institute of Water-Saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling 712100, Shaanxi Province, China; 3.Jiangsu Key Laboratory of Agricultural Meteorology, Nanjing University of Information Science and Technology, Nanjing 210044, China; 4.State Key Laboratory of Soil Erosion and Dryland Agriculture on the Loess Plateau, Institute of Water and Soil Conservation, Chinese Academy of Science and Ministry of Water Resource, Yangling 712100, Shaanxi Province, China)

Abstract: Surface solar radiation (R_s) data are important in hydrology, agriculture and ecology. Since only a few national meteorological stations have direct observation conditions, the Angström–Prescott (A–P) formula is widely used to estimate daily R_s . While the two empirical coefficients a and b required by the A–P formula have been recommended by the FAO (Food and Agriculture Organization) ($a = 0.25$; $b = 0.5$), recent studies have emphasized that the localization of the formula parameters could help to improve the estimation accuracy. This study used daily surface solar radiation (R_s) and other conventional meteorological data from 80 national weather stations with solar radiation observation data from 1967 to 2017 to derive reliable A–P formula coefficients in China mainland. First, the entire Chinese mainland was divided into four climatic zones: the Mountain Plateau Zone (MPZ), Subtropical Monsoon Zone (SMZ), Temperate Monsoon Zone (TMZ), and Temperate Continental Zone (TCZ). Next, the calibrated a and b values of the A–P formula were obtained at each weather station in different climate zones through linear regression, which were regarded as the observed values of the A–P formula coefficients. Four machine learning algorithms were applied to estimate the A–P formula coefficients. Each algorithm combined different input factor combinations to construct different estimation models for A–P formula coefficients. The accuracy of the estimated A–P formula coefficients and their impact on R_s estimation were evaluated. Some main conclusions were drawn as follows, when estimating the coefficient a , the SVM machine learning model with the five-factor input combination had the highest estimation accuracy, with $R^2 = 0.661$, $RMSE = 0.022$, and $nRMSE = 0.120$. When estimating the coefficient b , the ELM machine learning model with the four-factor input combination had the highest estimation accuracy, with $R^2 = 0.550$, $RMSE = 0.031$, and $nRMSE = 0.055$. Based on the A–P formula and the relevant coefficients a and b estimated with the selected optimal machine learning model to estimate R_s , the results showed that the $nRMSE$ of the machine learning model are 0.168, 0.225, 0.138, and 0.180 in the MPZ, SMZ, TMZ, and TCZ zones, respectively. Therefore, we recommended the SVM model with five-factor input combination to estimate the coefficient a , and the ELM model with four-factor input combination to estimate the coefficient b . Through which more accurate A–P formula coefficients could be obtained and thereby R_s estimation accuracy could be further improved with the A–P formula. This study provides a theoretical basis for the localization of A–P formula coefficients and the improvement of R_s estimation accuracy.

Key words: global solar radiation; Angström–Prescott; coefficient a and b ; machine learning

0 引言

太阳能是最重要的可再生能源之一，同时也是植物蒸腾地表蒸发的主要驱动力^[1]，在微观或宏观上推动了地球上的物理、生化和生物过程^[2]。准确获取太阳辐射数据，对于许多研究都有着至关重要的意义，例如作物模型^[3]、质量和能量转移评估^[4]、蒸散发建模^[5]，乃至光伏元件的设计^[6–8]等。在目前全球气候变化的大背景下，由于地表太阳总辐射决定了各区域农业生产的布局，影响粮食产量，所以越来越受到人们关注^[9–11]。

我国太阳能资源丰富^[12]，能否准确、高效地获取有关地表太阳总辐射的数据，对太阳能资源的合理利用和经济的可持续发展都具有十分重要的意义。然而，由于观测设备复杂、运行管理费、维护费过高等原因，导致地表太阳总辐射 (Surface Solar Radiation, R_s) 的观测相较于降水、气温、湿度、

日照时数等气象要素要困难许多^[13]。目前全国约有 2 500 多个国家气象站点，而其中具有直接太阳辐射观测的站点仅有 100 余个。因此，现有观测所得的太阳辐射数据已经远远不能满足太阳能生产与科研的需要^[14]，严重制约了太阳能的高效应用。为解决这个问题，需要根据已有的气象观测数据对太阳辐射数据进行间接估算。

国内外有许多学者致力于研究各种成本较低且精度较高的 R_s 估算方法。Meenal 和 Selvakuma^[15]研究表明估算 R_s 的输入因子中，纬度、最高温度、日照时数的影响最大，相对湿度的影响最小。Bristow 和 Campbell^[16]提出了近地面气温日较差和每日的太阳辐射值呈幂函数关系，并在此基础上建立了 B–C 模型，在此之后该模型被多次应用并被证明要优于很多的经验模型^[17–19]。目前，世界上应用最广泛的是基于日照时数的 Angström–Prescott 模型 (简称 A–P 模型)。该模型是由 Angström^[20]于 1924 年提出，由 Prescott^[21]于 1940 年在 Angström

公式基础上使用地外总辐射 R_a 替代晴空辐射改进而来。地外总辐射 R_a 和最大可能日照时数 N 的计算可参考联合国粮农组织 (Food and Agriculture Organization of the United Nations, FAO) 推荐的方法。Iziomon 和 Mayer^[22]发现应用 A-P 公式估算 R_s 的值与观测值的误差在 2.5%–3.4% 之间。Chen 等^[23]与 Liu 等^[24]也指出 A-P 公式精度更高。A-P 模型有两个经验系数 a 和 b , 它们的值往往需要根据研究区域当地的实测 R_s 数据来进行估算。但是由于可以观测 R_s 的气象站点较少, 通常无法直接获取特定区域的 a 和 b 系数校准值。当无法直接获得 a 和 b 系数的校准值时, FAO 建议系数 a 和 b 可以分别取 0.25 和 0.5。在国内, 鞠晓慧等^[25]提出中国大陆地区参数 a 的取值范围为 [0.1, 0.35]; 翁笃鸣^[26]与鞠晓慧等^[25]提出中国地区参数 b 的取值范围为 [0.35, 0.7]。可见这些建议参数值与 FAO 的推荐值直接存在较大差异, 并不能适应中国大陆所有的地区, 若是使用 FAO 推荐的 a 和 b 参数, 可能会影响 R_s 的估算精度。正如 Liu 等^[27]所总结的, 学者们一直在对 A-P 公式进行改进以提高其精度, 具体做法是将其参数与地理要素^[28]或气象要素^[29]相关联, 在方程中引入附加变量^[30]。然而 Liu 等^[27]研究表明修正后与修正前相比并没有明显的优势。因此, 反复的修正对提高其精度没有意义; 反而会使模型更加复杂, 增加数据收集难度。为了提高 A-P 模型的普适性, 将研究重点放在提高参数可用性上更有价值。

本研究分别采用不同的机器学习算法估算中国大陆地区不同气象站点的 A-P 公式经验系数 a 和 b 的值。主要包括: 气象站对应 A-P 公式系数校准值的获取; 基于机器学习算法的 A-P 公式系数估算; 利用 Angström-Prescott 模型和优化后 A-P 公式系数估算 R_s 。研究结果将为提高辐射估算精度提供一定的科学依据。

1 材料与方法

1.1 研究区域

本研究将中国大陆地区划分为高原山地气候区 (Mountain Plateau Zone, MPZ)、亚热带季风气候区 (Subtropical Monsoon Zone, SMZ)、温带季风气候区 (Temperate Monsoon Zone, TMZ)、温带大陆性气候区 (Temperate Continental Zone, TCZ)^[31], 其中 MPZ 地区海拔最高, 平均海拔为 4 236 m; 其次为 TCZ 地区, 平均海拔为 912 m; SMZ 地区平均海拔为 611 m, 海拔最低的 TMZ 地区平均海拔为 288 m。MPZ、SMZ、TMZ、TCZ 4 个气候区中分别有 7、34、23、16 个气象站点, 全国共有 80 个气象站点具有太阳辐射观测功能, 较为均匀地分布于中国大陆地区 (图 1)。在这 80 个气象站中采集的气象数据包括降雨量 (Precipitation, P , mm y^{-1})、相对湿度 (Relative Humidity, RH , %)、实际日照时数 (Actual Sunshine Hours, n , h)、潜在日照时数 (Potential Sunshine Hours, N , h)、平均温度 (Mean Temperature, T_{mean} , $^{\circ}C$)、最高温度 (Max Temperature, T_{max} , $^{\circ}C$)、最低温度 (Min Temperature, T_{min} , $^{\circ}C$)、风速 (Wind Speed, U , m/s)、地外太阳辐射 (Extraterrestrial Solar Radiation, R_a , MJ $m^{-2} d^{-1}$)、地表

太阳辐射 [Surface Solar Radiation, R_s , MJ/ ($m^2 \cdot d$)], 还包括各个站点的经度 (longitude, lon , rad)、纬度 (latitude, lat , rad) 以及高程 (altitude, alt , m)。此外还有 839 个不具有太阳辐射观测功能的一般气象站, 这些气象站只能采集上文中所述的一般气象数据。

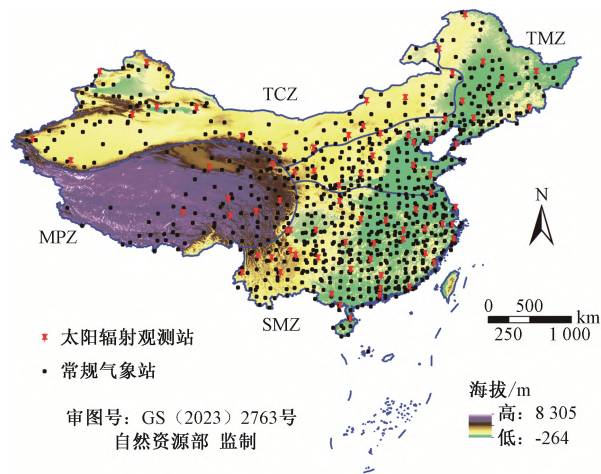


图 1 研究区示意图

Fig.1 Map of research area

1.2 Angström-Prescott 公式系数的校准

Angström-Prescott 公式是由 Angström^[20]于 1924 年提出, 由 Prescott^[21]于 1940 年在 Angström 公式基础上使用地外总辐射 R_a 替代晴空辐射改进而来 [公式组 (1)]。地外总辐射 R_a 和最大可能日照时数 N 的计算可参考 FAO 推荐的方法 [公式组 (1)]。

$$\begin{cases} R_s = R_a \left(a + b \frac{n}{N} \right) \\ R_a = (24 \times 60 / \pi) G_{sc} d_r (\omega_s \sin \varphi \sin \delta + \cos \varphi \cos \delta \sin \omega_s) \\ d_r = 1 + 0.033 \cos(2\pi \times J / 365) \\ \delta = 0.409 \sin(2\pi \times J / 365 - 1.39) \\ \omega_s = \arccos(-\tan \varphi \tan \delta) \\ N = 24 \times \omega_s / \pi \end{cases} \quad (1)$$

式中: R_s 为地表总辐射, MJ/ ($m^2 \cdot d$); R_a 为地外总辐射, MJ/ ($m^2 \cdot d$); a 、 b 为经验系数; n 为实际日照时数, h; N 为最大可能日照时数, h; G_{sc} 为太阳常数, 取 0.082 MJ/ ($m^2 \cdot min$); d_r 为日地间相对距离平方的倒数; ω_s 为太阳时角, rad; φ 为纬度, rad; δ 为太阳赤纬, rad; J 为年内某天的日序数。

本研究以 R_s/R_a 为因变量, n/N 为自变量, 建立线性方程式 (2), 然后运用最小二乘法进行线性回归拟合, 得到具有实测数据的全国 80 个气象站点的 A-P 公式系数校准值。

$$\frac{R_s}{R_a} = a + b \frac{n}{N} \quad (2)$$

式中符号意义同前。

1.3 输入气象因子的选择与组合

为了确定机器学习算法估算 A-P 公式系数时所需的输入气象因子组合, 本文根据课题组先前的研究结果, 从多个站点实测的气象预测因子中选取了与 R_s 相关性最高的 3 个因子: R_a 、 n/N 、 T_{max} ^[32], 这些预测因子是每个气象站点观测数据的多年平均值。刘峻明等^[33]在估算河南冬小麦产量时发现空间特征

在产量预测中非常重要，因此添加了2个地理因子纬度 lat 、海拔高度 alt ，这些预测因子是每个气象站点所在的地理位置。结合 A-P 公式中所包含的变量从这3个气象预测因子中挑出 R_a 、 n/N 作为基准因子，剩余因子与基准因子组合。最终得到了8种组合（表1）。由于三因子组合与四因子组合均有三种组合方式，因此后文精度评价表中只录入了各因子数精度最高的组合。

表1 输入因子组合表
Tab.1 The combination of input factors

序号	因子数	输入因子
1	2	$R_a, n/N$
2	3	$R_a + n/N + T_{\max}$
3		$R_a + n/N + lat$
4		$R_a + n/N + alt$
5	4	$R_a + n/N + T_{\max} + lat$
6		$R_a + n/N + T_{\max} + alt$
7		$R_a + n/N + lat + alt$
8	5	$R_a + n/N + T_{\max} + lat + alt$

1.4 A-P公式系数估算模型

1.4.1 支持向量机模型

支持向量机 (Support Vector Machines, SVM) 是一种具有结构简单、适应性强等优点的模型，它是一种有监督学习的机器学习算法，在处理非线性问题和小样本问题时具有较大的优势。顾万龙等^[34]应用支持向量机模型结合气象因子如气温和日照时数，建立了一个月尺度曝辐量的估算模型。研究表明，该方法可以在无辐射站点的情况下估算太阳辐射。此外，其他学者也将支持向量机应用于太阳辐射估算中^[35-37]。然而，目前关于使用支持向量机估算 A-P 公式系数的研究还相对较少。本研究使用 R 语言编程构建了 A-P 公式系数估算模型，核函数类型设置为“radial”。依据均方根误差 (Root mean square error, RMSE) 最小原则对参数 γ 和 C 进行寻优。

1.4.2 BP神经网络模型

BP 神经网络 (Back Propagation, BP) 是一种由输入层、隐含层和输出层组成的多层前馈神经网络，通过误差逆向传播算法进行训练。它具有很强的仿真性能，在许多非线性问题上具有良好的拟合性能，广泛应用于各个领域。周晋等^[38]应用神经网络模型结合气象数据资料建立了太阳辐射的估算模型，并与其他经验模型进行了对比。研究表明，基于神经网络的估算值与实测值更为吻合，且精度高于其他经验模型。然而，目前关于使用 BP 神经网络估算 A-P 公式系数的研究还相对较少。本研究使用 R 语言编程构建了 A-P 公式系数估算模型。依据均方根误差 (RMSE) 最小原则对隐藏节点数和权值衰减参数进行寻优，其余参数取默认值。

1.4.3 Cubist模型

Cubist 模型是由 Rule Quest 公司开发的一种决策树算法，它源于 M5 模型树^[39]，是一种改进的树模型^[40, 41]。Cubist 模型采用多个连续的分段式模型来预测未知量，因此具有以下优点：不仅可以解决线性和非线性问题，还可以提高模型的预

测精度。与简单的一元或多元线性回归相比，Cubist 模型的优势在于可以通过算法对输入空间进行划分。此外，Cubist 模型的训练时间短、效率高，并且能够处理具有高维属性的问题。本研究在训练前对自变量进行了归一化处理，然后使用 R 语言编程构建了 A-P 公式系数估算模型。依据均方根误差 (RMSE) 最小原则对参数 committees 和 neighbors 进行寻优。

1.4.4 极限学习机模型

极限学习机 (Extreme Learning Machine, ELM) 是一种基于前馈神经网络构建的机器学习算法，由输入层、隐含层和输出层组成，适用于监督学习和非监督学习问题。相比传统神经网络模型，ELM 具有训练速度更快、泛化性能更好的优点。已有学者将 ELM 应用于太阳辐射估算的研究中^[42]。本研究使用 R 语言编程构建了 A-P 公式系数估算模型。依据均方根误差 (RMSE) 最小原则对参数 nhid、actfun 和 init_weights 进行寻优。

1.5 统计指标

本研究采用决定系数、 R^2 、均方根误差、RMSE、归一化均方根误差、 $nRMSE$ 这3个统计评价指标来评价模型估算值与系数校准值之间的误差。 R^2 可以反映模型拟合精度的高低；RMSE 可以反映系数估算值与系数校准值之间的偏差程度； $nRMSE$ 是将均方根误差的表达式经过转换使其变为了无量纲的形式，是用来反映系数估算值与系数校准值之间偏差程度的归一化度量。当 R^2 越接近 1，RMSE 与 $nRMSE$ 越小，模型拟合精度越高。

$$R^2 = \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}}{\bar{y}} \quad (5)$$

式中： n 为样本数量； x_i 为 A-P 公式系数估算值； y_i 为 A-P 公式系数校准值； $\bar{x}$ 为 A-P 公式系数估算值的均值； $\bar{y}$ 为 A-P 公式系数校准值的均值。

2 结果与分析

2.1 A-P公式系数的校准

通过线性回归得到 80 个辐射测量站的 a 、 b 系数校准值（表2）。A-P 系数的平均值范围分别为 0.16~0.23 和 0.53~0.59。与 FAO 推荐值 ($a = 0.25$, $b = 0.50$) 相比，中国各区域的推荐系数值 a 可能较大，系数值 b 可能较小。系数 a 的校准值为 0.12~0.29，系数 b 的校准值为 0.48~0.72。在 MPZ 区域中，系数 a 和 b 的校准值均值均为最大，而在 SMZ 区域系数 a 的校准值均值最小，TCZ 区域系数 b 的校准值均值最小。

表2 中国不同气候区 Angström-Prescott (A-P)公式系数校准值的最大值、最小值、均值及标准差

Tab.2 Maximum (Max), minimum (Min), mean values, and standard errors (SE) of the benchmark values of coefficient of the Ångström-Prescott (A-P) formula in the four climatic zones in China

系数	气候区	站点数	最大值	最小值	均值±标准差
a	MPZ	7	0.29	0.20	0.23±0.03
	TCZ	16	0.27	0.18	0.22±0.02
	TMZ	23	0.28	0.14	0.19±0.03
	SMZ	34	0.21	0.12	0.16±0.02
	全国	80	0.29	0.12	0.19±0.04
b	MPZ	7	0.66	0.54	0.59±0.04
	TCZ	16	0.60	0.48	0.53±0.04
	TMZ	23	0.59	0.49	0.54±0.03
	SMZ	34	0.72	0.51	0.58±0.04
	全国	80	0.72	0.48	0.56±0.04

表3 基于SVM算法的A-P公式中参数a估算评价表

Tab.3 Evaluation of estimated parameter a in the A-P formula based on the SVM algorithm

序号	因子数	输入因子	训练			验证		
			R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
1	2	$R_a, n/N$	0.524	0.026	0.141	0.512	0.027	0.143
2	3	$R_a, n/N, alt$	0.646	0.022	0.119	0.643	0.023	0.121
3	4	$R_a, n/N, alt, T_{max}$	0.682	0.022	0.119	0.649	0.023	0.124
4	5	$R_a, n/N, alt, T_{max}, lat$	0.778	0.018	0.095	0.661	0.022	0.120

基于不同预测因子组合的BP模型在估算参数a时，五因子组合的估算能力最优，其模型训练与验证的RMSE、nRMSE是这5种组合里最小的，其模型训练RMSE为0.020，nRMSE为0.108，验证RMSE仅为0.023，nRMSE为0.124。双因子组合的估算精度最差，其模型训练与验证的R²与RMSE、nRMSE等评价指标均表明该因子组合在预测参数a时效果不如其余3

2.2 基于机器学习算法估算A-P公式系数

2.2.1 A-P公式系数a的估算

基于不同预测因子组合的SVM模型在估算参数a时，五因子组合的估算能力最优，其模型训练与验证的R²均大于其余所有因子组合，同时其RMSE、nRMSE是这5种组合里最小的，其模型训练RMSE仅为0.018，nRMSE为0.095，验证RMSE仅为0.022，nRMSE为0.120。双因子组合的估算精度最差，其模型训练与验证的R²与RMSE、nRMSE等评价指标均表明该因子组合在预测参数a时效果不如其余3种组合，其模型训练RMSE为0.026，nRMSE为0.141，验证RMSE为0.027，nRMSE为0.143（表3）。对于SVM模型来说，三因子组合相对于双因子组合加入了高程因子，精度得到了较大提高；四因子组合相对于三因子组合加入了气温因子，精度并没有显著性的提高，甚至四因子组合的验证nRMSE还增大了0.003；五因子组合相对于四因子组合加入了纬度因子，精度得到了较大提高。

表4 基于BP算法的A-P公式中参数a估算评价表

Tab.4 Evaluation of estimated parameter a in the A-P formula based on the BP algorithm

序号	因子数	因子	训练			验证		
			R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
1	2	$R_a, n/N$	0.472	0.027	0.145	0.424	0.029	0.154
2	3	$R_a, n/N, alt$	0.585	0.024	0.129	0.581	0.025	0.132
3	4	$R_a, n/N, alt, T_{max}$	0.701	0.020	0.109	0.625	0.023	0.124
4	5	$R_a, n/N, alt, T_{max}, lat$	0.705	0.020	0.108	0.624	0.023	0.124

基于不同预测因子组合的Cubist模型在估算参数a时，五因子组合的估算能力最优，其模型训练与验证的RMSE、nRMSE是这5种组合里最小的，其模型训练RMSE为0.020，nRMSE为0.107，验证RMSE仅为0.024，nRMSE为0.127。双因子组合的估算最差，其模型训练RMSE为0.023，nRMSE为0.127，验证RMSE为0.026，nRMSE为0.138（表5）。对于Cubist模型来说，三因子组合相对于双因子组合加入了高程因子，精度却没有显著性的提高；四因子组合相对于三因子组合加入了气温因子，精度得到了较大提高；五因子组合相对于四因子组合加入了纬度因子，精度却没有显著性的提高，甚至五因子组合模型训练R²还要低于四因子组合模型训

练R²。个组合，其模型训练RMSE为0.027，nRMSE为0.145，验证RMSE为0.029，nRMSE为0.154（表4）。对于BP模型来说，三因子组合相对于双因子组合加入了高程因子，精度得到了较大提高；四因子组合相对于三因子组合加入了气温因子，精度得到了较大提高；五因子组合相对于四因子组合加入了纬度因子，精度却没有显著性的提高。

基于不同预测因子组合的ELM模型在估算参数a时，五因子组合的估算能力最优，其模型训练与验证的RMSE、nRMSE是这5种组合里最小的，其模型训练RMSE为0.020，nRMSE为0.107，验证RMSE仅为0.023，nRMSE为0.124。双因子组合的估算精度最差，其模型训练RMSE为0.024，nRMSE为0.132，验证RMSE为0.030，nRMSE为0.160（表6）。对于ELM模型来说，三因子组合相对于双因子组合加入了高程因子，精度得到了较大提高；四因子组合相对于三因子组合加入了气温因子，精度却没有显著性的提高；五因子组合相对于四因子组合加入了纬度因子，精度得到了较大提高。

表 5 基于 Cubist 算法的 A-P 公式中参数 a 估算评价表
Tab.5 Evaluation of estimated parameter a in the A-P formula based on the Cubist algorithm

序号	因子数	因子	训练			验证		
			R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
1	2	$R_a, n/N$	0.599	0.023	0.127	0.544	0.026	0.138
2	3	$R_a, n/N, alt$	0.630	0.022	0.122	0.544	0.026	0.138
3	4	$R_a, n/N, alt, T_{max}$	0.718	0.020	0.107	0.610	0.024	0.127
4	5	$R_a, n/N, alt, T_{max}, lat$	0.710	0.020	0.109	0.614	0.024	0.127

表 6 基于 ELM 算法的 A-P 公式中参数 a 估算评价表
Tab.6 Evaluation table of estimated parameter a in the A-P formula based on the ELM algorithm

序号	因子数	因子	训练			验证		
			R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
1	2	$R_a, n/N$	0.563	0.024	0.132	0.456	0.030	0.160
2	3	$R_a, n/N, alt$	0.629	0.022	0.122	0.595	0.024	0.130
3	4	$R_a, n/N, alt, T_{max}$	0.652	0.022	0.118	0.590	0.024	0.131
4	5	$R_a, n/N, alt, T_{max}, lat$	0.712	0.020	0.107	0.622	0.023	0.124

使用 4 种机器学习模型分别结合不同的预测因子组合来估算系数 a ，整体来看所有模型对系数 a 最大值的估算都有不同程度的低估，所有模型的预测值里都未出现不合理的负值（图 2）。对于相同的机器学习算法，双因子组合的估算效果最差；五因子组合的估算效果最好，这与前文描述相符。对于相同的预测因子组合来说，双因子组合的最优算法是 Cubist 模型；三因子组合的最优算法

是 SVM 模型；四因子组合的最优算法是 SVM 模型；五因子组合的最优算法也是 SVM 模型。总之，SVM 模型的估算精度最优，ELM 模型的估算精度相对较差。所有模型里精度最低的是双因子组合的 ELM 模型，精度最高的是五因子组合的 SVM 模型，所以本研究将五因子组合的 SVM 模型作为机器学习算法估算 A-P 公式系数 a 的最优算法。

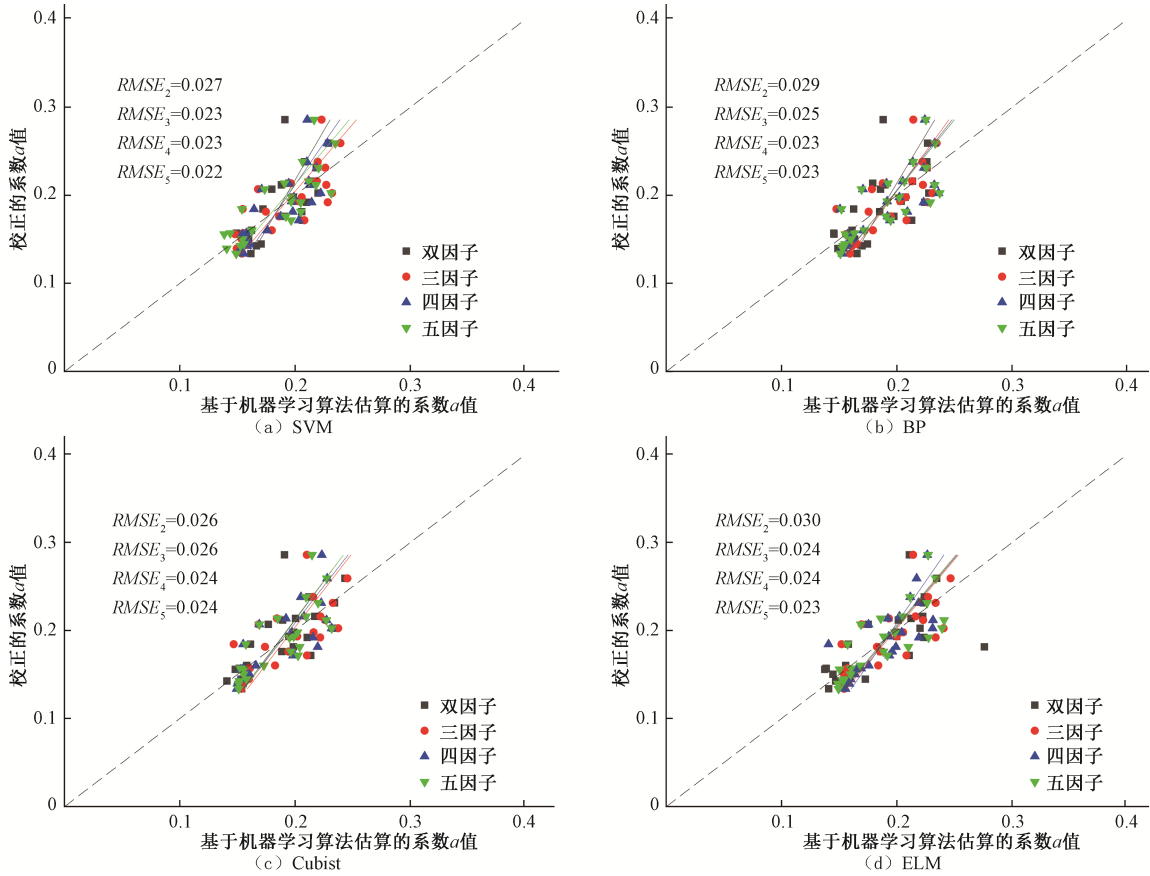


图 2 基于 4 种机器学习算法估算的 A-P 公式系数 a 与校准值之间的比较

Fig.2 Comparisons among the estimated coefficient a in the A-P formula based on the four different machine learning algorithms and the corresponding calibrated values

2.2.2 A-P公式系数 b 的估算

基于不同预测因子组合的SVM模型在估算参数 b 时,五因子组合的估算能力最优,其模型训练与验证的 $RMSE$ 、 $nRMSE$ 是这5种组合里最小的,其模型训练 $RMSE$ 仅为0.028, $nRMSE$ 为0.050,验证 $RMSE$ 为0.036, $nRMSE$ 为0.065。双因子组合的估算精度最差,其模型训练与验证的 R^2 与 $RMSE$ 、 $nRMSE$ 等评价指标均表明该因子组合在预测参数 a

时效果不如其余3个组合,其模型训练 $RMSE$ 为0.031, $nRMSE$ 为0.056,验证 $RMSE$ 为0.039, $nRMSE$ 为0.070(表7)。对于SVM模型来说,三因子组合相对于双因子组合加入了气温因子,精度得到了较大提高;四因子组合相对于三因子组合加入了高程因子,精度并没有显著性的提高;五因子组合相对于四因子组合加入了纬度因子,精度并没有显著性的提高。

表7 基于SVM算法的A-P公式中参数 b 估算评价表

Tab.7 Evaluation of estimated parameter b in the A-P formula based on the SVM algorithm

序号	因子数	因子	训练			验证		
			R^2	$RMSE$	$nRMSE$	R^2	$RMSE$	$nRMSE$
1	2	$R_a, n/N$	0.279	0.031	0.056	0.278	0.039	0.070
2	3	$R_a, n/N, T_{max}$	0.424	0.028	0.051	0.506	0.037	0.066
3	4	$R_a, n/N, T_{max}, alt$	0.443	0.029	0.052	0.515	0.036	0.065
4	5	$R_a, n/N, T_{max}, alt, lat$	0.472	0.028	0.050	0.491	0.036	0.065

基于不同预测因子组合的BP模型在估算参数 b 时,四因子组合的估算能力最优,其模型训练与验证的 $RMSE$ 、 $nRMSE$ 是这8种组合里最小的,其模型训练 $RMSE$ 为0.027, $nRMSE$ 为0.049,验证 $RMSE$ 为0.034, $nRMSE$ 为0.062。双因子组合的估算精度最差,其模型训练与验证的评价指标较差,其模型训练 $RMSE$ 为0.030, $nRMSE$ 为0.054,验证 $RMSE$ 为0.037,

$nRMSE$ 为0.066(表8)。对于BP模型来说,三因子组合相对于双因子组合加入了气温因子,精度得到了较大提高;四因子组合相对于三因子组合加入了纬度因子,精度没有显著性的提高;五因子组合相对于四因子组合加入了高程因子,精度没有显著性的提高,甚至验证 $RMSE$ 与 $nRMSE$ 分别都增大了0.001。

表8 基于BP算法的A-P公式中参数 b 估算评价表

Tab.8 Evaluation of estimated parameter b in the A-P formula based on the BP algorithm

序号	因子数	因子	训练			验证		
			R^2	$RMSE$	$nRMSE$	R^2	$RMSE$	$nRMSE$
1	2	$R_a, n/N$	0.313	0.030	0.054	0.358	0.037	0.066
2	3	$R_a, n/N, T_{max}$	0.427	0.027	0.050	0.459	0.034	0.062
3	4	$R_a, n/N, T_{max}, lat$	0.428	0.027	0.049	0.460	0.034	0.062
4	5	$R_a, n/N, T_{max}, lat, alt$	0.442	0.027	0.049	0.447	0.035	0.063

基于不同预测因子组合的Cubist模型在估算参数 b 时,四因子组合的估算能力最优,其模型训练与验证的 $RMSE$ 、 $nRMSE$ 是这5种组合里最小的,其模型训练 $RMSE$ 为0.023, $nRMSE$ 为0.041,验证 $RMSE$ 仅为0.032, $nRMSE$ 为0.057。双因子组合的估算精度最差,其模型训练 $RMSE$ 为0.026,

$nRMSE$ 为0.046,验证 $RMSE$ 为0.035, $nRMSE$ 为0.062(表9)。对于Cubist模型来说,三因子组合相对于双因子组合加入了气温因子,精度得到了提高;四因子组合相对于三因子组合加入了纬度因子,精度没有显著性的提高;五因子组合相对于四因子组合加入了高程因子,精度没有提高反而低于四因子组合。

表9 基于Cubist算法的A-P公式中参数 b 估算评价表

Tab.9 Evaluation of estimated parameter b in the A-P formula based on the Cubist algorithm

序号	因子数	因子	训练			验证		
			R^2	$RMSE$	$nRMSE$	R^2	$RMSE$	$nRMSE$
1	2	$R_a, n/N$	0.517	0.026	0.046	0.463	0.035	0.062
2	3	$R_a, n/N, T_{max}$	0.574	0.024	0.043	0.474	0.033	0.059
3	4	$R_a, n/N, T_{max}, lat$	0.602	0.023	0.041	0.510	0.032	0.057
4	5	$R_a, n/N, T_{max}, lat, alt$	0.587	0.023	0.042	0.467	0.033	0.060

基于不同预测因子组合的ELM模型在估算参数 b 时,四因子组合的估算能力最优,其模型训练与验证的 $RMSE$ 、 $nRMSE$ 是这8种组合里最小的,其模型训练 $RMSE$ 为0.022, $nRMSE$ 为0.040,验证 $RMSE$ 仅为0.031, $nRMSE$ 为0.055。双因子组合的估算精度最差,其模型训练 $RMSE$ 为0.029, $nRMSE$ 为0.052,验证 $RMSE$ 为0.036, $nRMSE$ 为0.064(表10)。对于ELM模型来说,三因子组合相对于双因子组合加入了气温因

子,精度得到了较大提高;四因子组合相对于三因子组合加入了纬度因子,精度却没有显著性的提高;五因子组合相对于四因子组合加入了高程因子,精度反而降低了许多。

使用4种机器学习模型分别结合不同的预测因子组合估算系数 b ,整体来看所有模型对系数 b 最大值的估算都有不同程度的低估,但所有模型的系数预测值里都未出现不合理的负值(图3)。BP、Cubist、ELM这3种模型都出现了五因子组合

表 10 基于 ELM 算法的 A-P 公式中参数 b 估算评价表
Tab.10 Evaluation of estimated parameter b in the A-P formula based on the ELM algorithm

序号	因子数	因子	训练			验证		
			R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
1	2	$R_a, n/N$	0.357	0.029	0.052	0.458	0.036	0.064
2	3	$R_a, n/N, T_{\max}$	0.632	0.022	0.040	0.530	0.031	0.055
3	4	$R_a, n/N, T_{\max}, lat$	0.622	0.022	0.040	0.550	0.031	0.055
4	5	$R_a, n/N, T_{\max}, lat, alt$	0.444	0.027	0.049	0.424	0.036	0.064

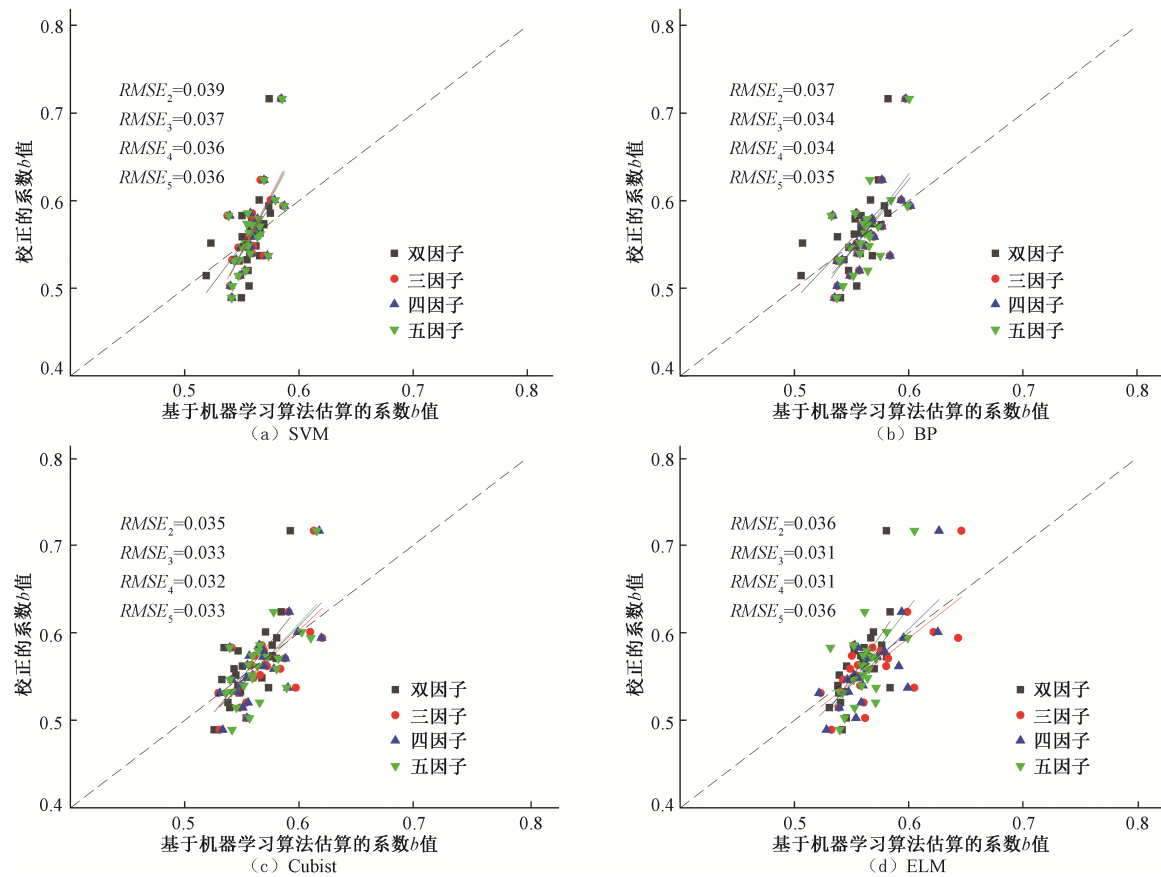


图 3 基于 4 种机器学习算法估算的 A-P 公式系数 b 与校准值之间的比较

Fig.3 Comparisons among the estimated coefficient b in the A-P formula based on the four different machine learning algorithms and the corresponding calibrated values

的精度低于四因子组合的情况，这可能是由于系数 b 的复杂性或是系数 b 与预测因子间的相关性所导致的。对于相同的机器学习算法，总体来看双因子组合的估算效果最差，四因子组合的估算效果最好的，这与前文描述相符。对于相同的预测因子组合来说，双因子组合的最优模型是 Cubist 模型；三因子组合的最优模型是 ELM 模型；四因子组合的最优模型是 ELM 模型；五因子组合的最优模型是 Cubist 模型。总之，ELM 模型的估算精度最优，SVM 模型的估算精度相对较差，所有模型里精度最低的是双因子组合的 SVM 模型，精度最高的是四因子组合的 ELM 模型。所以，本研究将四因子组合的 ELM 模型作为机器学习算法估算 A-P 公式系数 b 的最优算法。

2.3 基于估算 A-P 公式系数的地表太阳辐射估算

挑选出机器学习算法中估算精度最高的模型及对应输入因子组合，将其估算的系数 a 、 b 代入 A-P 公式，联合各个站点多年逐日气象数据共同驱动 A-P 公式，得到 R_i 的逐日估算

值，并与各个站点多年逐日观测值相比较（图 4）。可以发现，TMZ 地区的估算精度最高， $nRMSE$ 为 0.138；而 MPZ 地区的估算精度最低，虽然其 $nRMSE$ 值只比 TMZ 地区大，但其数据点回归线与 1:1 线之间还存在一定误差，这可能是由于 MPZ 地区的太阳辐射值偏大从而导致 $nRMSE$ 值相对于 SMZ、TCZ 地区较小。SMZ、TCZ 两个地区的估算精度次于 TMZ 地区。在全国范围内都存在不同程度的将辐射极小值偏大估算的情况，其中最严重的是 SMZ 地区；同时 TCZ、TMZ 地区存在对极大值存在不同程度的偏小估算现象。但与系数估算精度相比，驱动 A-P 公式估算 R_i 的精度要更高，因为其回归线与 1:1 线之间误差更小。

3 讨论

对中国各地 A-P 公式系数的估算方法进行研究，对于提升辐射估算的准确性和实现太阳能资源的有效利用具有极其

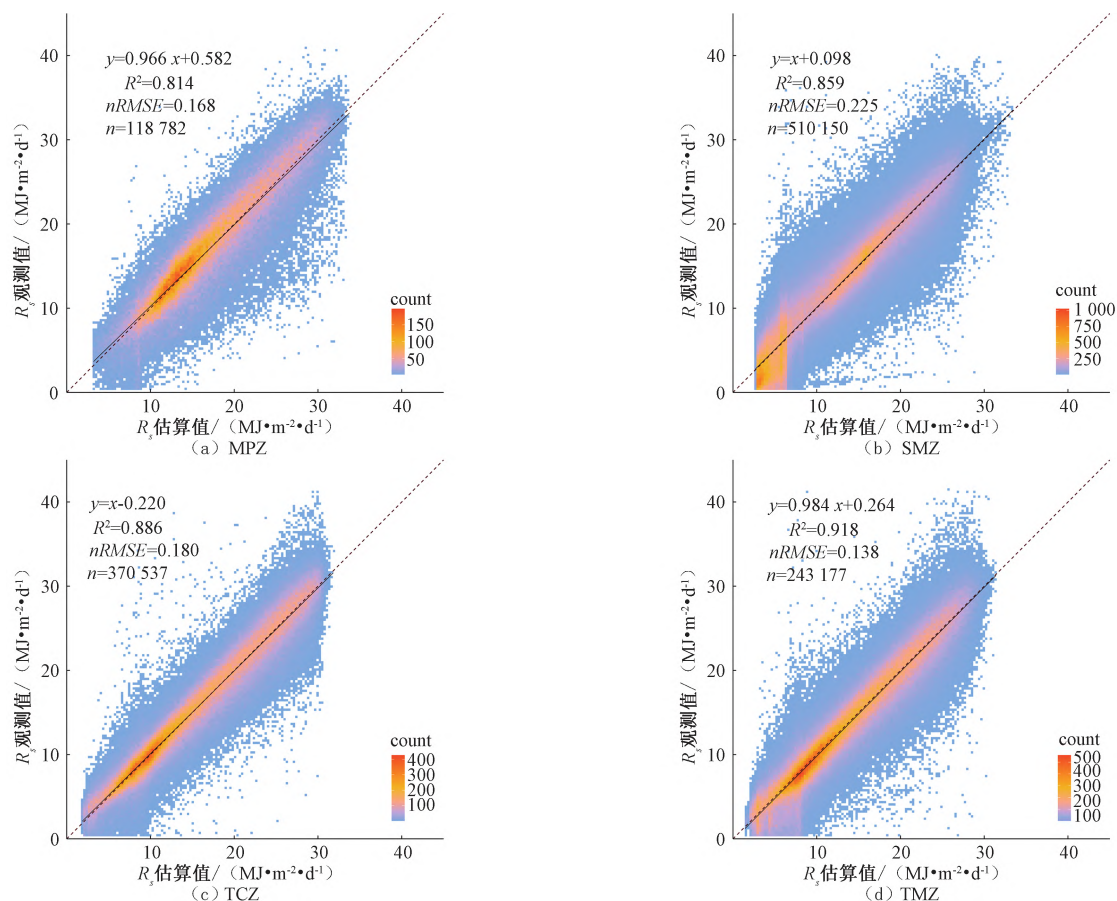


图4 机器学习算法估算的 a 、 b 系数驱动A-P公式估算的 R_s 与实测值之间的比较

Fig.4 Estimated R_s based on the A-P formula and related a and b coefficients estimated through machine learning models.

关键的作用。A-P公式中系数 a 的物理意义是当阴天时（即 $n=0$ 时）地外辐射 R_a 中可以到达地球表面（即 R_s ）的比例，而 $a+b$ 的物理意义是当晴天时（即 $n>0$ 时）地外辐射 R_a 中可以到达地球表面（即 R_s ）的比例。而SMZ地区由于降雨丰沛，因此一年中阴天较多，这会造成 R_s 观测值较低，而 R_a 只与日地距离、日序数等时空参数有关因此会导致SMZ地区系数 a 校准值与FAO推荐值（ $a=0.25$ ； $b=0.50$ ）存在极大误差，在晴天由于系数 a 偏小因此系数 b 在校准时会偏大，因此 $a+b$ 基本与FAO推荐值相差不大。其余地区也有不同程度的偏差，其中MPZ地区 $a+b$ 值最大，这可能是由于该地区海拔较高，太阳能资源丰富而阴天较少，导致系数 a 小于FAO推荐值的同时 $a+b$ 值远大于FAO推荐值。这说明了在使用A-P系数估算 R_s 之前进行系数本地化是十分重要的，否则可能会造成较大的估算误差。

在使用机器学习算法估算A-P系数 a 时，随着输入因子的增加，估算精度不断提高，但估算精度明显提高的转折点多在双因子至三因子以及四因子至五因子这两个节点，结合表3~表6中数据可以发现双因子至三因子精度明显提高是因为加入了高程；而四因子至五因子精度提高是因为加入了纬度。这表明在估算A-P系数 a 时空间特征可以作为一个重要因素。刘峻明等^[33]在估算河南冬小麦产量时发现空间特征在产量预测中非常重要，这也同样说明了在类似情况中可以首要考虑将地理因子作为估算变量之一。而在估算系数 b 时并没有表现出如同系数 a 的规律，随着输入因子的增加，估算精度并没有不

断提高，五因子组合的估算精度还要低于四因子组合。而估算精度明显提高的转折点在双因子至三因子这个节点，结合表7~表10)中数据发现精度明显提高是因为加入了最高温度。这可能是由于系数 b 的复杂性或是系数 b 与预测因子间的相关性所导致的。机器学习算法在绝大多数站点系数估算中都可以较为准确的估算出对应站点的系数值，在个别站点中估算值远小于校准值。这可能是由于机器学习算法自身还存在缺陷，无法对极大极小值做出准确预测，也有可能是在系数校准中个别站点观测数据太少，导致在最小二乘法拟合时该站点系数校准值自身存在较大误差。

在使用机器学习估算的系数值驱动A-P公式估算 R_s 时不同地区的 R_s 估算精度不同，同时A-P公式在全国范围内都对辐射极小值有着不同程度的低估，其中以SMZ地区最为明显，这可能是由于SMZ为中国南方地区，水资源丰富，因为降雨较为频繁所以导致观测数据存在误差，导致估算存在高估现象。而在TCZ和TMZ地区对极大值存在不同程度的低估，这可能是由于TCZ、TMZ地区为中国北方地区，该地区拥有大面积但实测站点数却过少，导致仅有站点无法完全代表其特征，在A-P系数校准时便存在误差，导致最后在驱动A-P公式估算 R_s 时对极大值存在不同程度的低估。

A-P公式系数本地化在提高 R_s 估算精度与估算简易程度上具有重要意义，但第一步系数校准时需要大量的实测数据，这使得各个缺乏辐射实测数据的站点难以通过参数校准实现参数本地化，基于不同气象、地理因子估算A-P公式系数则

可以解决这个问题。此外还有学者在估算 A-P 公式系数时没有直接估算系数 a 与 b ，而是采取两步法来进行计算^[43-45]，即先估算系数 a 和 $a+b$ 或系数 b 和 $a+b$ ，接着在此基础上求差值从而得到系数 b 或系数 a 。在本研究中机器学习算法虽然取得了较良好的估算精度，但并未使用两步法，而是直接基于气象、地理因子估算系数 a 和 b ，因此在应用机器学习算法估算 A-P 系数时应用两步法究竟可以得到更好的结果或是由于增加估算流程使得误差传递的更大仍是一个值得进一步探讨的问题。

4 结 论

(1) 在全国范围内，联合国粮农组织推荐值 ($a = 0.25$; $b = 0.50$) 高估了系数 a (0.19 ± 0.04) 并低估了系数 b (0.56 ± 0.04) 校准值。

(2) 所有模型对 a 、 b 系数最大值的估算都存在不同程度的低估。估算系数 a 时，精度最高的是五因子组合的 SVM 模型，所以推荐应用 SVM 模型与 $R_a + n/N + alt + T_{\max} + lat$ 输入因子组合来估算 A-P 公式系数 a 。估算系数 b 时，精度最高的是四因子组合的 ELM 模型，所以推荐应用 ELM 模型与 $R_a + n/N + T_{\max} + lat$ 输入因子组合来估算 A-P 公式系数 b 。

(3) 在全国 4 个不同气候区的 R_s 估算时，A-P 公式在不同地区存在低估或高估的现象，但经过误差传递后 R_s 估算精度要优于系数估算精度。所以仍然推荐 A-P 公式作为通用方法来估算缺测站点的太阳辐射值。

参考文献:

- [1] SCHULZE R E. A physically based method of estimating solar radiation from suncards[J]. *Agricultural Meteorology*, 1976, 16: 85-101.
- [2] ALMOROX J, BOCCO M, WILLINGTON E. Estimation of daily global solar radiation from measured temperatures at Canada de Luque, Cordoba, Argentina[J]. *Renewable Energy*, 2013, 60: 382-387.
- [3] BANSOULEH B F, SHARIFI M A, VAN KEULEN H. Sensitivity analysis of performance of crop growth simulation models to daily solar radiation estimation methods in Iran[J]. *Energy Conversion and Management*, 2009, 50(11): 2 826-2 836.
- [4] NIKOLOV N, ZELLER K F. Modeling coupled interactions of carbon, water, and ozone exchange between terrestrial ecosystems and the atmosphere. I: Model description[J]. *Environmental Pollution*, 2003, 124(2): 231-246.
- [5] VAZQUEZ R F, FEYEN J. Potential evapotranspiration for the distributed modeling of Belgian catchments[J]. *Journal of Irrigation and Drainage Engineering*, 2004, 130(1): 1-8.
- [6] GHONEIM A A. Design optimization of photovoltaic powered water pumping systems[J]. *Energy Conversion and Management*, 2006, 47 (11-12): 1 449-1 463.
- [7] ISMAIL M S, MOGHAVVEMI M, MAHLIA T M I. Analysis and evaluation of various aspects of solar radiation in the Palestinian territories[J]. *Energy Conversion and Management*, 2013, 73: 57-68.
- [8] FERNANDEZ-PACHECO D G, MOLINA-MARTINEZ J M, RUIZ-CANALES A, et al. A new mobile application for maintenance tasks in photovoltaic installations by using GPS data[J]. *Energy Conversion and Management*, 2012, 57: 79-85.
- [9] BESHARAT F, DEGHAN A A, FAGHIH A R. Empirical models for estimating global solar radiation: A review and case study[J]. *Renewable & Sustainable Energy Reviews*, 2013, 21: 798-821.
- [10] PINKER R T, ZHANG B, DUTTON E G. Do satellites detect trends in surface solar radiation?[J]. *Science*, 2005, 308(5 723): 850-854.
- [11] SINGH G K. Solar power generation by PV (photovoltaic) technology: A review[J]. *Energy*, 2013, 53: 1-13.
- [12] HE G, KAMMEN D M. Where, when and how much solar is available? A provincial-scale solar resource assessment for China [J]. *Renewable Energy*, 2016, 85: 74-82.
- [13] THORNTON P E, RUNNING S W. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation[J]. *Agricultural and Forest Meteorology*, 1999, 93(4): 211-228.
- [14] PAN T, WU S, DAI E, et al. Estimating the daily global solar radiation spatial distribution from diurnal temperature ranges over the Tibetan Plateau in China[J]. *Applied Energy*, 2013, 107: 384-393.
- [15] MEENAL R, SELVAKUMAR A I. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters[J]. *Renewable Energy*, 2018, 121: 324-343.
- [16] BRISTOW K L, CAMPBELL G S. On the relationship between incoming solar-radiation and daily maximum and minimum temperature[J]. *Agricultural and Forest Meteorology*, 1984(31): 159-166.
- [17] LIU D L, SCOTT B J. Estimation of solar radiation in Australia from rainfall and temperature observations[J]. *Agricultural and Forest Meteorology*, 2001, 106(1): 41-59.
- [18] THORNTON P E, RUNNING S W, WHITE M A. Generating surfaces of daily meteorological variables over large regions of complex terrain[J]. *Journal of Hydrology*, 1997, 190(3-4): 214-251.
- [19] LIU X, MEI X, LI Y, et al. Evaluation of temperature-based global solar radiation models in China[J]. *Agricultural and Forest Meteorology*, 2009, 149(9): 1 433-1 446.
- [20] ÅNGSTRÖM A. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation[J]. *Quarterly Journal of the Royal Meteorological Society*, 1924, 50(210): 121-126.
- [21] PRESCOTT J A. Evaporation from a water surface in relation to solar radiation[J]. *Transactions of the Royal Society of South Australia*, 1940, 46: 114-118.

- [22] IZIOMON M G, MAYER H. Assessment of some global solar radiation parameterizations[J]. *Journal of Atmospheric and Solar-Terrestrial Physics*, 2002, 64(15): 1 631-1 643.
- [23] CHEN R S, ERSI K, YANG J P, et al. Validation of five global radiation models with measured daily data in China[J]. *Energy Conversion and Management*, 2004, 45(11-12): 1 759-1 769.
- [24] LIU J, LIU J, LINDERHOLM H W, et al. Observation and calculation of the solar radiation on the Tibetan Plateau[J]. *Energy Conversion and Management*, 2012, 57: 23-32.
- [25] 鞠晓慧, 屠其璞, 李庆祥. 我国太阳总辐射气候学计算方法的再讨论[J]. *南京气象学院学报*. 2005(4): 516-521. JU X H, TU Q P, LI Q X, Discussion on the climatological calculation of solar radiation[J]. *Journal of Nanjing Institute of Meteorology*, 2005(4): 516-521.
- [26] 翁笃鸣. 试论总辐射的气候学计算方法[J]. *气象学报*, 1964(3): 304-315. Ун Ду - мин, О КЛИМАТОЛОГИЧЕСКИХ МЕТОДАХ РАСЧЕТА СУММАРНОЙ РАДИАЦИИ[J]. *Acta Meteorologica Sinica*, 1964(3): 304-315.
- [27] LIU X, MEI X, LI Y, et al. Calibration of the angstrom-prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin[J]. *Agricultural and Forest Meteorology*, 2009, 149(3-4): 697-710.
- [28] ELAGIB N A, MANSELL M G. New approaches for estimating global solar radiation across Sudan[J]. *Energy Conversion and Management*, 2000, 41(5): 419-434.
- [29] HALOUANI N, NGUYEN C T, VONGOC D. Calculation of monthly average global solar-radiation on horizontal surfaces using daily hours of bright sunshine[J]. *Solar Energy*, 1993, 50(3): 247-258.
- [30] ODOJO J C, SULAIMAN A T, AIDAN J, et al. The importance of maximum air-temperature in the parameterization of solar-radiation in Nigeria[J]. *Renewable Energy*, 1995, 6(7): 751-763.
- [31] SONG Y, ACHBERGER C, LINDERHOLM H W. Rain-season trends in precipitation and their effect in different climate regions of China during 1961-2008[J]. *Environmental Research Letters*, 2011, 6(3).
- [32] 何 川. 基于机器学习的太阳辐射和参考作物腾发量估算研究[D]. 陕西杨凌: 西北农林科技大学, 2020. HE C, Estimation of solar radiation and reference crop evapotranspiration based on machine learning Methods[D]. Yangling, Shaanxi: Northwest A & F University, 2020.
- [33] 刘峻明, 和晓彤, 王鹏新, 等. 长时间序列气象数据结合随机森林法早期预测冬小麦产量[J]. *农业工程学报*. 2019, 35(6): 158-166. LIU J M, HE X T, WANG P X, et al. Early prediction of winter wheat yield with long time series meteorological data and random forest method[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(6): 158-166.
- [34] 顾万龙, 朱业玉, 潘 攀, 等. 支持向量机方法在太阳辐射计算中的应用[J]. *太阳能学报*. 2010, 31(1): 56-60. GU W L, ZHU Y Y, PAN P, et al. Application of SVM methods to calculation of solar radiation[J]. *Acta Energaie Solaris Sinica*, 2010, 31(1): 56-60.
- [35] 杨 光, 谭言科, 吴 凡. 支持向量机方法在南京太阳总辐射推算中的应用[J]. *大气科学学报*. 2017, 40(5): 708-714. YANG G, TAN Y K, WU F, Application of support vector machine method in estimation of total solar radiation in Nanjing[J]. *Trans Atmos Sci*, 2017, 40(5): 708-714.
- [36] 彭怀午, 杨晓峰, 刘方锐. 基于 SVM 方法的太阳总辐射值的估算[J]. *电网与清洁能源*, 2010, 26(12): 82-84. PENG H W, YANG X F, LIU F R, Estimating global solar radiation based on SVM method[J]. *Power System and Clean Energy*, 2010, 26(12): 82-84.
- [37] 李 净, 温松楠. 基于 3 种机器学习法的太阳辐射模拟研究[J]. *遥感技术与应用*. 2020, 35(3): 615-622. LI J, WEN S N, Simulation of Solar Radiation based on Three Machine Learning Method[J]. *Remote Sensing Technology and Application*, 2020, 35 (3): 615-622.
- [38] 周 晋, 吴业正, 晏 刚, 等. 利用神经网络估算太阳辐射[J]. *太阳能学报*. 2005(4): 509-512. ZHOU J, WU Y Z, YAN G, et al. solar radiation estimation using artificial neural networks[J]. *Acta Energaie Solaris Sinica*, 2005(4): 509-512.
- [39] 戴 舒, 付迎春, 赵耀龙. 基于 Cubist 模型树的都市不透水面百分比遥感估算模型[J]. *地球信息科学学报*. 2016, 18(10): 1 399-1 409. DAI S, FU Y C, ZHAO Y L. The remote sensing model for estimating urban impervious surface percentage based on the cubist model tree[J]. *Journal of Geo-information Science*, 2016, 18(10): 1 399-1 409.
- [40] GRAY J M, BISHOP T F A, WILFORD J R. Lithology and soil relationships for soil modelling and mapping[J]. *Catena*, 2016, 147: 429-440.
- [41] FRANK E, WANG Y, INGLIS S, et al. Technical note: using model trees for classification[J]. *Machine Learning*, 1998, 32(1): 63-76.
- [42] SALCEDO-SANZ S, CASANOVA-MATEO C, PASTOR-SANCHEZ A, et al. Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization - Extreme Learning Machine approach[J]. *Solar Energy*, 2014, 105: 91-98.
- [43] HE C, LIU J, XU F, et al. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods[J]. *Energy Conversion and Management*, 2020, 220.
- [44] LIU X, MEI X, LI Y, et al. Variation in reference crop evapotranspiration caused by the Angstrom-Preseott coefficient: Locally calibrated versus the FAO recommended[J]. *Agricultural Water Management*, 2009, 96(7): 1 137-1 145.
- [45] LIU X, LI Y, ZHONG X, et al. Towards increasing availability of the Angstrom-Preseott radiation parameters across China: Spatial trend and modeling[J]. *Energy Conversion and Management*, 2014, 87: 975-989.