

冯 瑞,杨丽萍,侯成磊,等. 基于随机森林的陕西省西安市近地表气温估算[J]. 地球科学与环境学报, 2022, 44(1): 102-113.
FENG Rui, YANG Li-ping, HOU Cheng-lei, et al. Estimation of Near-surface Air Temperature in Xi'an City of Shaanxi Province, China Based on Random Forest[J]. Journal of Earth Sciences and Environment, 2022, 44(1): 102-113.

DOI: 10.19814/j.jese.2021.08027

基于随机森林的陕西省西安市近地表气温估算

冯 瑞¹, 杨丽萍^{2*}, 侯成磊³, 王 彤², 张 静², 肖 舜⁴

(1. 长安大学 地球科学与资源学院, 陕西 西安 710054; 2. 长安大学 地质工程与测绘学院, 陕西 西安 710054; 3. 山东农业工程学院 国土资源与测绘工程学院, 山东 济南 250100; 4. 陕西师范大学 地理科学与旅游学院, 陕西 西安 710119)

摘 要:随着城市规模不断扩大以及人口激增,城市气候与热环境问题日益凸显,开展城市近地表气温遥感监测研究能够为改善城市气候、减缓热岛效应、打造适宜人居环境提供参考。针对传统气温监测方法在多因素复杂关系模拟中的局限性,以陕西省西安市为研究对象,运用可以集成多要素、学习复杂、非线性映射关系的随机森林(Random Forest, RF)模型,基于 Landsat 8 卫星遥感数据以及 SRTM 高程数据相关参数的综合分析,构建多种近地表气温估算的随机森林模型,通过性能对比评估优选最佳模型,估算了 2016 年 5 月 16 日西安市近地表气温,分析了近地表气温的空间分布特征。结果表明:在所有近地表气温影响因子中,高程对随机森林模型近地表气温估算的贡献度最大,其次是地表温度。所有随机森林模型训练集的判定系数(R^2)均高于 0.916,均方根误差(RMSE)均低于 0.467 °C,验证集判定系数均高于 0.726,均方根误差均低于 0.840 °C;训练集判定系数均高于验证集,均方根误差均低于验证集;最优随机森林模型训练集判定系数为 0.934,均方根误差为 0.425 °C,验证集判定系数为 0.795,均方根误差为 0.783 °C;气温估算精度判定系数为 0.792,均方根误差为 1.055 °C。西安市中心城区气温高于郊县区,中心城区最低气温平均值、最高气温平均值及气温平均值分别高于郊县区 1.54 °C、0.01 °C 和 1.76 °C。综上所述,西安市近地表气温南低北高,空间差异明显,自中心城区、郊县区至南部山区逐渐降低,呈现出显著的城市热岛效应。

关键词:近地表气温;随机森林;估算;地表温度;热岛效应;高程;Landsat 8;西安

中图分类号:P407;TP79

文献标志码:A

文章编号:1672-6561(2022)01-0102-12

Estimation of Near-surface Air Temperature in Xi'an City of Shaanxi Province, China Based on Random Forest

FENG Rui¹, YANG Li-ping^{2*}, HOU Cheng-lei³, WANG Tong²,
ZHANG Jing², XIAO Shun⁴

(1. School of Earth Science and Resources, Chang'an University, Xi'an 710054, Shaanxi, China; 2. School of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, Shaanxi, China; 3. College of Land Resources and Surveying & Mapping Engineering, Shandong Agriculture and Engineering University, Jinan 250100, Shandong, China; 4. School of Geography and Tourism, Shaanxi Normal University, Xi'an 710119, Shaanxi, China)

收稿日期:2021-08-16;修回日期:2021-11-03 投稿网址: <http://jese.chd.edu.cn/>

基金项目:国家自然科学基金项目(41371220,42071345);陕西省重点研发计划项目(2020ZDLSF06-07);中央高校基本科研业务费专项资金项目(300102269112)

作者简介:冯 瑞(1996-),女,陕西延安人,理学硕士研究生,E-mail:fengruichd@163.com.

* 通讯作者:杨丽萍(1968-),女,陕西铜川人,副教授,理学博士,E-mail:zylpyang@chd.edu.cn.

Abstract: With the continuously expanding of the urban areas and the dramatic population growth, serious urban climate and thermal environment issues have occurred and attracted extensive attention. High-precision and large-scale remote sensing monitoring of near-surface air temperature in Xi'an city can provide reference for the improvement of urban climate, the alleviation of heat island effect, as well as the creation of a suitable living environment. Traditional near-surface air temperature monitoring approaches are confined when dealing with complicated relationships among multiple factors. Taking Xi'an city as the study area, the random forest (RF) models, which can integrate multiple factors and simulate complicated and non-linear mapping relationship, were adopted. Through comprehensive analysis of various parameters derived from Landsat 8 imagery and SRTM DEM data, several near-surface air temperature estimation RF models were constructed, and the optimal model was confirmed by performance comparison and evaluation. Furthermore, near-surface air temperature in Xi'an city on May 16, 2016 was estimated, and the spatial distribution pattern was discussed. The results show that among all the near-surface air temperature influencing factors, altitude has the highest contribution to the estimated temperature of the RF models, followed by land surface temperature (LST); both of them are crucial for the estimation of near-surface air temperature. For all the RF models, the training set R^2 is higher than 0.916, and the RMSE is lower than 0.467 °C; all the validation set R^2 is higher than 0.726, and the RMSE is lower than 0.840 °C; the training set R^2 of all RF models is higher than that of the validation set, while the training set RMSE is lower compared with the validation set; the training set R^2 of the optimal model is 0.934 with the RMSE being 0.425 °C, and the validation set R^2 is 0.795 with a RMSE value of 0.783 °C; the R^2 and RMSE of the air temperature estimation accuracy are 0.792 and 1.055 °C, respectively. Near-surface air temperature in the central urban districts of Xi'an city is higher than that of suburban areas, and the mean minimum temperature, the mean maximum temperature, as well as the mean temperature are higher than those of suburban areas by 1.54 °C, 0.01 °C and 1.76 °C, respectively. The near-surface air temperature in the study area is low in the south and high in the north, and presents obvious spatial heterogeneity. Air temperature gradually decreases from the central urban districts through suburbs to the southern mountainous areas, indicating a prominent urban heat island effect.

Key words: near-surface air temperature; random forest; estimation; land surface temperature; heat island effect; altitude; Landsat 8; Xi'an

0 引 言

近地表气温指距离地面 1.5~2.0 m 的大气温度。作为描述地表大气环境的重要指标,近地表气温是气象观测的基本要素之一,也是地表与大气能量交换、水循环及生态过程研究的重要输入因子^[1-4]。近地表气温与人类的日常生活和社会生产活动息息相关,高精度、大范围近地表气温监测对于生态环境、农业生产、气候变化等研究具有重要的现实意义^[5-8]。

气象站点观测数据是获取近地表气温的直接途径,具有准确性高等特点。早有学者基于气象观测

数据,利用反距离权重、样条函数和克里金插值等多种空间插值法获取了具有空间连续性的气温数据^[9-11]。然而受多种因素限制,通常情况下气象站点数量有限,在站点稀疏和分布不均匀的地区,利用空间插值法得到的气温数据误差较大^[12-13]。相较于传统的观测方法,卫星遥感技术具有大范围、实时性和经济性等优势,同时可提供比气象站点观测资料更完整的空间异质度信息^[14]。随着卫星遥感技术不断发展,国内外学者基于遥感数据在近地表气温反演中开展了大量研究,提出了多种近地表气温反演算法,主要可概括为以下 4 类。第一类为常规统计法,包括单因子统计法和多因子统计法。该方法

通常基于地表温度 (Land Surface Temperature, LST) 与实际观测气温数据间的相关关系进行气温模拟。陈命男利用一元线性回归方法分别对上海市 2007 年和 2008 年的地表温度与气温观测数据进行拟合, 判定系数 (R^2) 分别为 0.566 0 和 0.706 6^[15]; Lin 等在结合归一化植被指数 (Normalized Difference Vegetation Index, NDVI) 和环境绝对湿度等因子估算复杂地形山区气温时发现, 判定系数从 0.60~0.65 增加到 0.71~0.88, 这表明在解释复杂地形表面能量通量的时空变化时必须考虑相关的环境因素^[16]; Mohammadi 等利用地理加权回归方法比较了单变量和多变量模型反演大气温度的精度, 发现采用多变量回归方法精度更高, 均方根误差 (Root Mean Square Error, RMSE) 为 0.62 °C, 平均相关系数为 0.99^[17]。第二类为温度植被指数 (Temperature-Vegetation Index, TVX) 法。该方法是一种利用地表温度和光谱植被指数间的负相关性, 从遥感数据中提取气温的空间邻域运算方法^[1,18], 适用于浓密植被覆盖地区。徐永明等通过去除温度植被指数空间窗口内残余水体和云像元改进温度植被指数法, 提高了算法适用范围^[19]。第三类为能量平衡法。Pape 等基于能量平衡方程, 并结合下垫面植被及地貌数据反演近地表气温, 均方根误差为 0.37 °C~1.02 °C^[20]。第四类为机器学习法。Yoo 等设计了 8 种不同变量输入方案, 利用随机森林算法模型研究了洛杉矶市和首尔市的日最高气温和最低气温, 最佳模型判定系数大于 0.7, 均方根误差小于 1.7 °C^[21]; 高亮等采用随机森林、支持向量机、AdaBoost 和岭回归等 4 种机器学习模型拟合观测气温与影响要素的相关关系, 结果表明随机森林模型精度最高, 判定系数平均值为 0.85, 均方根误差平均值为 0.50 °C^[22]; 邢立亭等使用随机森林模型估算了兰州市近地表最高气温和最低气温, 模型估算效果良好, 判定系数分别为 0.921 和 0.916^[23]。

综上所述, 前人已提出了多种近地表气温遥感反演方法, 并在近地表气温监测中取得了显著成效。其中, 因考虑了多种环境因子, 多因子统计法的反演精度通常高于单因子统计法; 温度植被指数法输入参数少且相对简单, 但对于低植被及裸土地区, 该方法并不适用; 能量平衡法具有深厚的物理基础, 但计算过程相对繁杂; 机器学习法已广泛应用于各个领域, 已有学者利用机器学习法在近地表气温反演中进行了有益尝试, 大量研究表明随机森林模型在获取高时空分辨率近地表气温中精度较高, 表现良

好^[21-23]。气温的时空分布与变化受海拔高度、地形、下垫面性质和纬度等多种因素的综合影响^[24], 且影响关系复杂。传统的统计回归模型对这一复杂耦合关系的表达能力有限; 随机森林模型训练速度快, 模型参数调整相对简单, 可以较好地拟合因变量与多个自变量的非线性关系, 为探讨这一复杂耦合关系提供了新的途径。

陕西省西安市是中国西部地区的中心城市。近年来, 随着西安市向国际化大都市的快速迈进, 城市化进程飞速发展, 人口激增, 城市规模不断扩张, 人工景观大幅代替自然景观^[25]。人类密集活动以及城市下垫面性质的变化加剧了西安市局地气候与环境问题。前人研究表明, 西安市平均气温呈上升趋势, 且市区气温上升速率高于郊县区^[26-27]。以城市热岛效应为典型的城市气候与热环境问题愈演愈烈, 这一问题不仅可以通过热危害直接威胁人体健康, 还会加重空气污染, 间接增加呼吸系统和心脑血管等疾病的发病率^[28]。在上述背景下, 开展西安市近地表气温估算研究, 以期改善城市气候、减缓城市热岛效应、打造适宜人居环境提供参考。

本文首先基于西安市 Landsat 8 卫星数据提取地表温度、归一化植被指数、归一化建筑指数 (Normalized Difference Built-up Index, NDBI)、改进的归一化水体指数 (Modified Normalized Difference Water Index, MNDWI) 和地表反照率 (Albedo) 等 5 种遥感因子, 基于 30 m 空间分辨率的 SRTM DEM 数据提取高程 (Altitude)、坡向 (Aspect) 和坡度 (Slope) 等 3 种地形因子, 将提取的 8 个参数进行相关性和重要性综合分析; 然后在此基础上进行不同参数组合, 构建多个估算气温的随机森林模型; 最后基于气象站点观测数据, 利用交叉验证的方法评估模型性能、验证模型精度, 并选取最优参数组合方案进行近地表气温估算。

1 研究区概况及数据来源

1.1 研究区概况

西安市地处陕西省关中平原中部, 经度范围为 107.40°E~109.49°E, 纬度范围为 33.42°N~34.45°N (图 1)。西安市整体呈南高北低、阶梯状的地势特点, 城市以北的渭河平原和南部的秦岭山地形成强烈的地貌对照。西安市管辖 11 个区和 2 个县, 其中灞桥区、未央区、新城区、莲湖区、雁塔区、碑林区为西安市中心城区。长安区、临潼区、高陵区、阎良区、鄠邑区、蓝田县和周至县是西安市的周边郊县区。山区主要指东西向横跨周至县、鄠邑区、长安

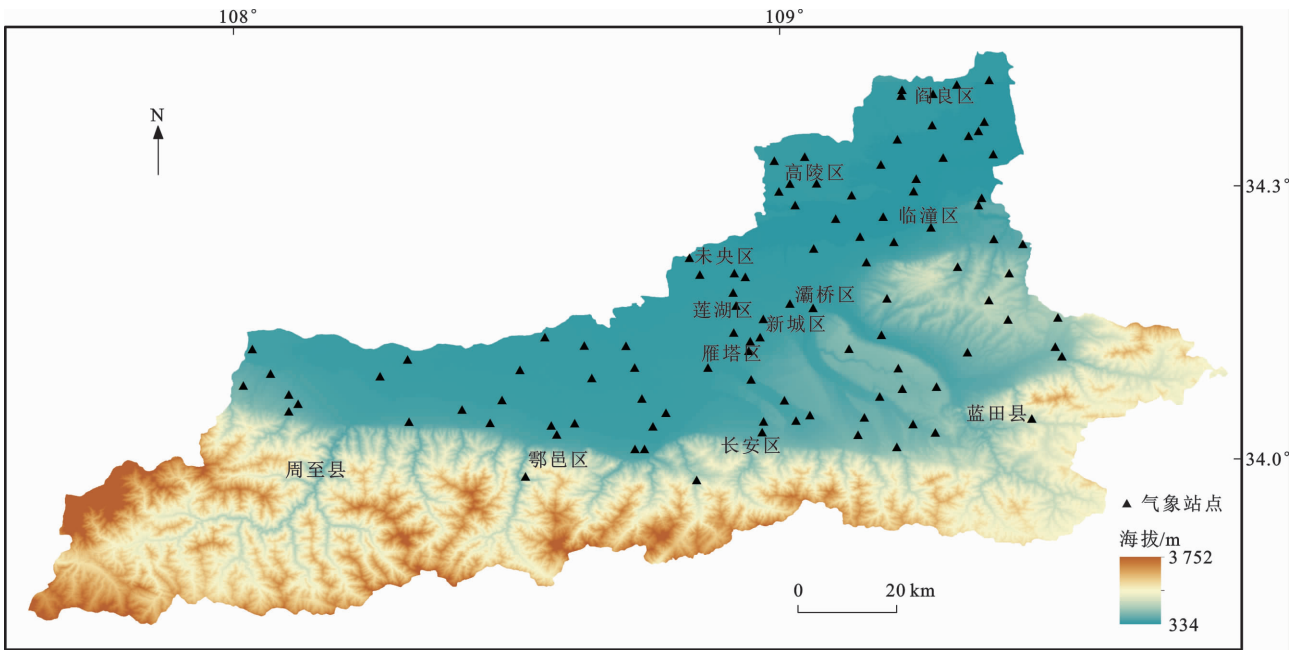


图 1 陕西省西安市地区气象站点分布

Fig. 1 Distribution of Meteorological Stations in Xi'an City of Shaanxi Province

区和蓝田县的秦岭山区。西安市属暖温带半湿润大陆性季风气候,四季分明,年平均气温为 13.0℃~13.7℃,年降水量为 522.4~719.5 mm。

1.2 数据来源

遥感数据采用 2016 年 5 月 16 日的 Landsat 8 卫星影像,数据来源于地理空间数据云 (<http://www.gscloud.cn/>)。利用 ENVI5.3 软件对卫星影像进行了辐射定标、大气校正及拼接与裁剪。以 30 m 空间分辨率的 SRTM DEM 作为高程数据,数据来源于美国地质调查局 (<http://www.usgs.gov/>),将 6 幅 DEM 数据进行拼接、裁剪得到覆盖全研究区的 DEM 数据。

气象数据为 2016 年 5 月 16 日西安市地面自动气象站点逐小时气象观测数据,采用与卫星过境时间相一致的观测数据进行模型性能评估与精度验证。研究区共 102 个气象站点,由于卫星过境时不是整点时刻,所以取过境时相邻整点实测气温数据进行线性插值计算出卫星过境时的气温^[29]。研究区气象站点分布如图 1 所示。

2 研究方法

2.1 自变量获取

2.1.1 地表温度

本文采用胡德勇等针对 Landsat 8 第 10 波段 (TIRS 10)提出的地表温度单窗算法(简称为 TIRS 10-SC 算法)^[30]计算地表温度。胡德勇等利用 TI-

RS 10-SC 算法、覃志豪单窗算法和 Jiménez-Muñoz 算法反演了不同植被覆盖类型的地表温度,认为 TIRS 10-SC 算法可以较好地应用于 Landsat 8 卫星影像数据地表温度反演^[31]。TIRS 10-SC 算法公式为

$$T = \frac{[K_1(\varphi_1 + \varphi_2)T_{10} + (1 - \varphi_1 - \varphi_2)T_{10}^2 - K_1\varphi_2T_a]}{K_1\varphi_1} \tag{1}$$

$$\varphi_1 = \epsilon_{10}\tau_{10} \tag{2}$$

$$\varphi_2 = (1 - \tau_{10}) \cdot [1 + (1 - \epsilon_{10})\tau_{10}] \tag{3}$$

式中: T 为地表温度; T_{10} 为 TIRS 10 的亮温; T_a 为大气平均作用温度; ϵ_{10} 为 TIRS 10 地表比辐射率; τ_{10} 为 TIRS 10 大气透射率; K_1 取常数 1 321.08 K; φ_1 和 φ_2 为中间参量。

2.1.2 其他自变量

影响近地表气温的因素众多,除地表温度外,太阳辐射、植被、水体、建筑物分布及地形等因素均具有一定影响。因此,本文引入其他遥感因子(包括归一化植被指数、归一化建筑指数、改进的归一化水体指数和地表反照率)以及地形因子(包括高程、坡向和坡度)作为构建随机森林气温估算模型的输入参数。遥感因子基于 Landsat 8 卫星数据获取,计算公式见表 1,地形因子由 DEM 数据获取。

2.2 随机森林

随机森林是 Breiman 提出的一种基于决策树组合的可用于分类、回归以及多维数据处理的机器学习

表 1 遥感因子提取算法

Table 1 Extraction Algorithms for Remote Sensing Factors

遥感因子	公式	参考文献
归一化 植被指数	$I_{NDVI} = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$	[32]
改进的归一 化水体指数	$I_{MNDWI} = \frac{\rho_{GREEN} - \rho_{MIR}}{\rho_{GREEN} + \rho_{MIR}}$	[33]
归一化 建筑指数	$I_{NDBI} = \frac{\rho_{MIR} - \rho_{NIR}}{\rho_{MIR} + \rho_{NIR}}$	[34]
地表反照率	$I_{Albedo} = \frac{(0.356\rho_{BLUE} + 0.130\rho_{RED} + 0.373\rho_{NIR} + 0.085\rho_{SWIR1} + 0.072\rho_{SWIR2} - 0.0018) - \alpha_p}{T_{sw}};$ $T_{sw} = 0.75 + 2 \times 10^{-5} Z$	[35]

注： I_{NDVI} 为归一化植被指数； I_{MNDWI} 为改进的归一化水体指数； I_{NDBI} 为归一化建筑指数； I_{Albedo} 为地表反照率； ρ_{NIR} 、 ρ_{RED} 、 ρ_{GREEN} 、 ρ_{MIR} 、 ρ_{BLUE} 、 ρ_{SWIR1} 和 ρ_{SWIR2} 分别为Landsat 8的近红外、红光、绿光、中红外、蓝光、短波红外1和短波红外2波段的反射率； T_{sw} 为大气单向透过率； Z 为海拔高度； α_p 为大气辐射值影响，通常取值0.03。

习算法^[36]。较传统决策树算法而言,该算法能够平衡分布不均匀样本的误差,分类精度高,对异常值和噪声有很好的容忍度^[37]。随机森林是未经修剪的分类树或回归树的集合,这些树是通过训练数据的自举样本和树归纳中的随机特征选择而创建的,随机森林选取总样本数的 2/3 作为训练集构建决策树,剩余 1/3 数据用于验证所构建的模型性能^[38]。通常可通过调整决策树的数量以及单棵决策树的特征数量来提高随机森林模型的性能,本文使用循环迭代法进行参数优化。对于决策树数量设置,将决策树数量步长设置为 10,自第 10 棵树起每加入 10 棵树模型运行 1 次,决策树数量至 300 棵树时模型结束,将判定系数最大值对应的模型作为此次运算的最佳结果保存。对于单棵决策树数量的设置,在 1 到模型输入参数数量之间选择一个合适的值作为单棵决策树的特征数量,通过在决策树棵树的循环下嵌套一个循环体,将步长设置为 1,把表现最好的作为单棵决策树的特征数量进行保存。本文随机森林模型的构建通过 MATLAB 软件中的 Random Forest 工具箱完成,模型实现步骤如下。

步骤一:通过自举采样方法 Bootstrap 从原始数据集中随机抽取 M 次(有放回),样本集组成 M 个训练集,未被抽到的数据为袋外(Out of Bag, OOB)数据,可用于评估构建模型的性能^[39]。

步骤二:新构建的每个训练集均单独作为一棵决策树,每棵决策树在生长过程中均不进行剪枝。

步骤三:重复步骤二 n 次,将生成的 n 棵决策树组成随机森林,并应用于分类和回归。对于每棵树都能得到一个袋外数据误差统计,对所有决策树的袋外数据误差取平均后可得到随机森林泛化误差估

计^[40],袋外数据误差是无偏估计。

2.3 变量重要性计算

随机森林可以在训练过程中生成输入变量的重要性度量。本文随机森林中的变量重要性利用平均精度下降(Mean Decrease Accuracy,MDA)进行表征,该方法是基于袋外数据误差的一种重要性评判方法。其基本原理是在外包样本数据集中改变某一输入参数值,然后计算造成的估算误差,通过得到的误差评判参数的重要性。若误差较大,说明该变量对预测结果越敏感,同时重要性值越大,对模型的贡献越高。

2.4 模型验证方法

随机森林是随机选取样本数据的过程,其本身具有交叉验证的优势,当模型中决策树的数目足够保证每个样本都可以充当一次训练集和样本集时,可以最大程度地避免模型过度拟合,同时提升模型的外推能力^[41]。K 折交叉验证是对交叉验证的进一步发展和推广^[42],根据 Kohavi 的研究结果^[43],10 倍的变异系数在估计模型预测误差时效果最佳,因此,本文采用 10 折交叉验证来评估模型性能,采用判定系数和均方根误差作为模型性能的评价指标。

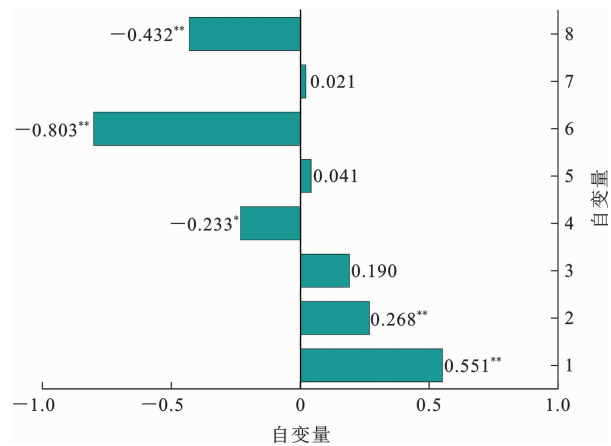
3 结果分析

3.1 变量相关性与重要性分析

3.1.1 相关性分析

为定量分析自变量与近地表气温之间的关系,本文选用 Pearson 相关系数对近地表气温和 8 个输入参数之间的潜在关系进行研究,结果如图 2 所示。

相关性涉及的 8 个自变量包括 5 个遥感因子和 3 个地形因子。遥感因子中,地表温度、改进的归一



** 表示在 0.01 水平(双侧)上显著相关; * 表示在 0.05 水平(双侧)上显著相关。纵坐标中,1 表示地表温度;2 表示改进的归一化水体指数;3 表示归一化建筑指数;4 表示归一化植被指数;5 表示地表反照率;6 表示高程;7 表示坡向;8 表示坡度

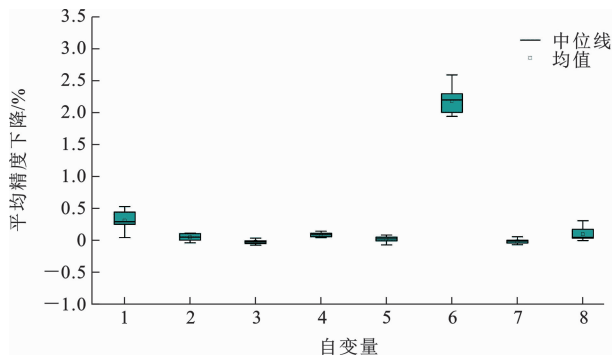
图 2 相关性分析结果

Fig. 2 Correlation Analysis Results

化水体指数和归一化植被指数通过了显著性检验,归一化建筑指数和地表反照率未通过显著性检验。其中,地表温度相关性最高,相关系数为 0.551;地表反照率相关性最低,相关系数为 0.041。地形因子中,高程和坡度通过显著性检验,坡向未通过显著性检验。高程相关性最高,相关系数为-0.803;坡向相关性最低,相关系数仅 0.021。整体来看,8 个自变量的相关性从高到低依次为高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数、归一化建筑指数、地表反照率、坡向。其中,地表温度、改进的归一化水体指数、归一化建筑指数、地表反照率和坡向与近地表气温成正相关关系;归一化植被指数、高程和坡度与近地表气温成负相关关系;与近地表气温相关性最高的参数为高程,相关性最低的为坡向。

3.1.2 重要性分析

输入因子重要性评分用平均精度下降表示,结果如图 3 所示。由图 3 可见:在遥感因子中,地表温度对于近地表气温估算的重要性高于其他 4 个因子,地表温度的重要性评分最高,所对应的平均值也最高;归一化建筑指数重要性评分与平均值最低;改进的归一化水体指数、归一化植被指数和地表反照率的重要性评分以及平均值差别不大,表明这 3 个变量对于近地表气温估算的贡献度相当。在地形因子中,高程对于近地表气温估算的重要性评分最高,坡度次之,坡向最低,其对应的平均值亦然。整体来看,8 个自变量中,高程的重要性评分及平均值最



横坐标中,1 表示地表温度;2 表示改进的归一化水体指数;3 表示归一化建筑指数;4 表示归一化植被指数;5 表示地表反照率;6 表示高程;7 表示坡向;8 表示坡度

图 3 重要性分析结果

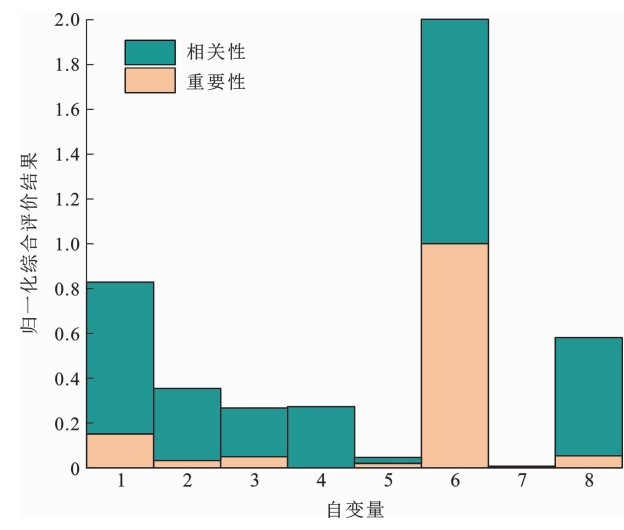
Fig. 3 Importance Analysis Results

高,地表温度次之,其他 6 个自变量的重要性评分没有明显差异,归一化建筑指数重要性评分及平均值最低。从重要性分析来看,对近地表气温估算贡献度最大的是高程,其次是地表温度。

3.1.3 相关性与重要性综合分析

第 3.1.1 和 3.1.2 节分别通过自变量与实测气温的相关性和重要性分析了影响因子对近地表气温估算的贡献度以及相关性和重要性表征变量不同的特征意义。这两种评价指标具有不同的量纲,为了综合考虑自变量的相关性和重要性,并消除不同指标之间的量纲影响,对相关系数和重要性评分进行归一化(MIN-MAX Normalization)处理,将数据结果映射到[0,1],为判定自变量的贡献度大小及设计不同模型参数组合方案奠定基础。

归一化处理结果如图 4 所示。由图 4 可见:在遥感因子中,地表温度归一化结果最高,这与地表温度和近地表气温的相关性和重要性较高相一致,其次为改进的归一化水体指数,归一化结果最低的参数是地表反照率,表明遥感因子中对近地表气温估算贡献度最大的是地表温度;在地形因子中,高程的归一化结果最高,这与地形因子高程和近地表气温的相关性和重要性最高相一致,其次为坡度,最低为坡向,表明高程对估算近地表气温贡献度最大。整体来看,在构建随机森林模型估算近地表气温的过程中,8 个自变量的贡献度从大到小依次为高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数、归一化建筑指数、地表反照率、坡向,表明在近地表气温估算过程中地表温度和高程是最为重要的两个输入参数。近地表气温变化受多因素影响,一般情况下,在对流层中其垂直分布随高程的增加而



横坐标中,1 表示地表温度;2 表示改进的归一化水体指数;
3 表示归一化建筑指数;4 表示归一化植被指数;5 表示地表
反照率;6 表示高程;7 表示坡向;8 表示坡度

图 4 归一化综合评价结果

Fig. 4 Normalized Comprehensive Assessment Results

降低,主要是由于距离地表越远,吸收的地面长波辐射越少,气温越低。地表吸收太阳辐射而增温,然后向外发出长波辐射,近地表气温因吸收地面辐射而增温,地表热量对近地表气温变化起直接的、主导的作用,因此,地表温度与近地表气温相关性较高。归一化植被指数反映植被覆盖度状况,植被覆盖度的变化会改变地表潜热、感热通量以及植被自身的蒸腾作用,进而对近地表气温变化产生影响,在近地表气温估算时考虑地表热量、水体、建筑等下垫面状况是非常有必要的。因此,根据归一化综合评价结果所反映的影响因子对近地表气温的贡献度大小,构建 P2~P8(其中数字代表参数个数)共 7 种参数组合方案(表 2)。

3.2 随机森林近地表气温估算模型对比与性能评估

将第 3.1.3 节构建的 7 种参数组合方案分别输入随机森林模型进行近地表气温估算,采用 10 折交叉验证评估模型性能,结果如表 3 所示。由表 3 可知:7 种参数组合方案下的随机森林模型训练集判定系数均高于 0.916,均方根误差均低于 0.467℃;所有模型验证集判定系数均高于 0.726,均方根误差均低于 0.840℃;训练集判定系数均高于验证集,均方根误差均低于验证集。7 种参数组合方案分别构建的随机森林模型均能够对训练集和验证集进行较好的拟合,表明随机森林在研究区近地表气温估算中具有较好效果。

进一步分析发现,对于训练集,P5~P8 方案模型判定系数大于 P2~P4 方案模型,且其均方根误

差小于 P2~P4 方案模型,随着输入参数的增加,P5~P8 方案模型判定系数和均方根误差无明显变化,模型性能较 P2~P4 方案模型表现较好且趋于稳定。对于验证集,P2~P5 方案模型判定系数随着参数的增加而增加,均方根误差随着参数的增加而减小,P5 方案模型性能表现较好;相比 P2~P5 方案模型,P6~P8 方案模型判定系数出现先减后增变化,均方根误差出现先增后减变化。综合训练集和验证集来看,当全部参数输入模型后,随机森林模型(P8 方案模型)性能表现最佳,其验证集精度优于 P5 方案模型,P8 方案模型的训练集判定系数为 0.934,均方根误差为 0.425℃(P5 方案模型训练集判定系数为 0.935,均方根误差为 0.426℃),验证集判定系数为 0.795,均方根误差为 0.783℃(P5 方案模型验证集判定系数为 0.788,均方根误差为 0.785℃)。因此,本文基于 P8 方案建立研究区随机森林近地表气温估算模型。

3.3 结果验证

从上述随机森林模型估算近地表气温的训练集及验证集结果来看,随机森林近地表气温估算模型精度较高,其中 P8 方案模型精度最高。为进一步验证模型精度,对基于最佳模型的近地表气温估算值与站点观测值进行线性拟合,结果如图 5 所示。由图 5 可见,近地表气温估算值与站点观测值的判定系数为 0.792,均方根误差为 1.055℃,气温估算精度较好。根据样本点相对 1:1 线的分布情况可知,部分近地表气温估算值出现高值低估的现象。这一方面可能是由于近地表气温观测站点几乎全部分布在秦岭以北地区,研究区南部气温观测站点较为有限(图 1),所以观测气温的高值较多,在模型估算近地表气温过程中出现了一定程度的过拟合;另一方面,在气温估算过程中,未能考虑太阳辐射对地表温度直接作用的影响,从而也会导致近地表气温高值低估的情况。综上所述,虽然出现部分高值低估的现象,但根据总体估算精度来看,低估偏差在合理范围内,所构建的随机森林模型适用于西安市近地表气温的估算。

3.4 西安市近地表气温估算与空间分布特征

基于上述最佳随机森林模型,模拟得到 2016 年 5 月 16 日西安市近地表气温(图 6)。由图 6 可见,西安市近地表气温呈现显著的空间差异性,整体呈南低北高、中心城区高于郊县区的态势。

为进一步分析近地表气温的空间差异性,提取

表 2 模型参数组合方案

Table 2 Combination Schemes of Model Parameter

参数组合方案	参数个数	参数名称
P2	2	高程、地表温度
P3	3	高程、地表温度、坡度
P4	4	高程、地表温度、坡度、改进的归一化水体指数
P5	5	高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数
P6	6	高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数、归一化建筑指数
P7	7	高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数、归一化建筑指数、地表反照率
P8	8	高程、地表温度、坡度、改进的归一化水体指数、归一化植被指数、归一化建筑指数、地表反照率、坡向

表 3 随机森林模型性能

Table 3 Model Performance of Random Forest

参数组 合方案	训练集		验证集	
	判定系数	均方根误差/℃	判定系数	均方根误差/℃
P2	0.916	0.459	0.726	0.840
P3	0.922	0.447	0.756	0.792
P4	0.921	0.467	0.776	0.791
P5	0.935	0.426	0.788	0.785
P6	0.934	0.424	0.768	0.805
P7	0.936	0.428	0.772	0.802
P8	0.934	0.425	0.795	0.783

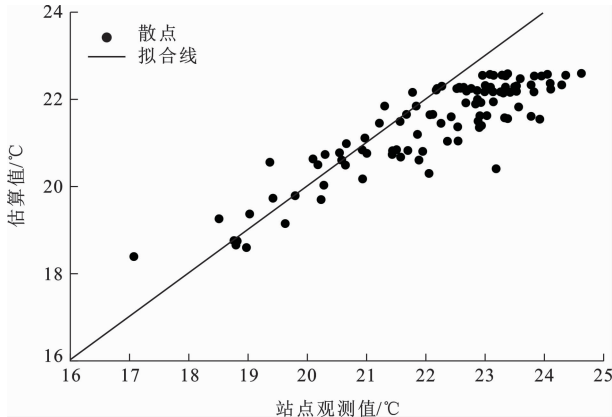


图 5 近地表气温估算值和站点观测值散点图

Fig. 5 Scatter Plot of Estimated Temperature and Site Temperature

了不同区县的最低气温、最高气温、平均气温和平均高程,结果如表 4 所示。由表 4 可见,西安市 13 个区县的最低气温平均值为 20.50℃,最高气温平均值为 23.49℃,并且各区县最高气温均高于 23.00℃,西安市平均气温为 21.98℃。中心城区的 3 个气温指标均大于郊县区,其最低气温平均值、最高气温平均值以及气温平均值分别高于郊县区 1.54℃、0.01℃和 1.76℃。13 个区县平均高程为 380~1 433 m,从平均高程来看,秦岭山区横跨的周至县、鄠邑区、长安区和蓝田县的平均高程远高于其他辖区,这是引起这 4 个区县平均气温普遍低于其

表 4 气温-高程分区统计

Table 4 Statistics of Air Temperature-altitude in Different Districts

参数		最低气温/ ℃	最高气温/ ℃	平均气温/ ℃	平均高程/ m
中心 城区	灞桥区	19.20	23.51	22.19	484
	未央区	21.69	23.46	23.06	381
	莲湖区	21.98	23.46	23.19	402
	雁塔区	20.73	23.51	22.81	436
	碑林区	22.31	23.51	23.14	416
	新城区	22.10	23.51	23.17	417
	平均值	21.33	23.49	22.93	423
郊县区	高陵区	21.67	23.46	22.74	381
	阎良区	21.58	23.44	22.61	380
	临潼区	19.07	23.50	21.81	485
	蓝田县	19.05	23.45	19.81	1 101
	长安区	19.05	23.51	20.71	995
	鄠邑区	19.05	23.51	20.63	1 080
	周至县	19.05	23.51	19.86	1 433
	平均值	19.79	23.48	21.17	836
西安市整体		20.50	23.49	21.98	645

他辖区的一个重要原因,对于高程相近的其他辖区,近地表气温仍表现出空间差异性。伴随着城市化发展,地表下垫面类型发生改变,尤其对于中心城区,高层建筑、道路铺装等均会使植被、土地和水域等自然表面变为不透水面,进而改变了地表长波辐射能量,导致气温发生变化,这是城市气温产生空间差异的另一原因。此外,近地表气温受植被覆盖度、水域面积、人口密度及社会生产活动等众多因素影响,与中心城区相比,郊县区发展较慢且区域面积大,人类社会生产活动相对较少,使郊县区平均气温普遍低于中心城区。

在中心城区内,灞桥区具有最低气温(19.20℃),远低于其他 5 个辖区,差值为 1.53℃~3.11℃;灞桥区平均气温为 22.19℃,低于除雁塔区以外的其他 4 个辖区,差值为 0.62℃~1.00℃。灞桥

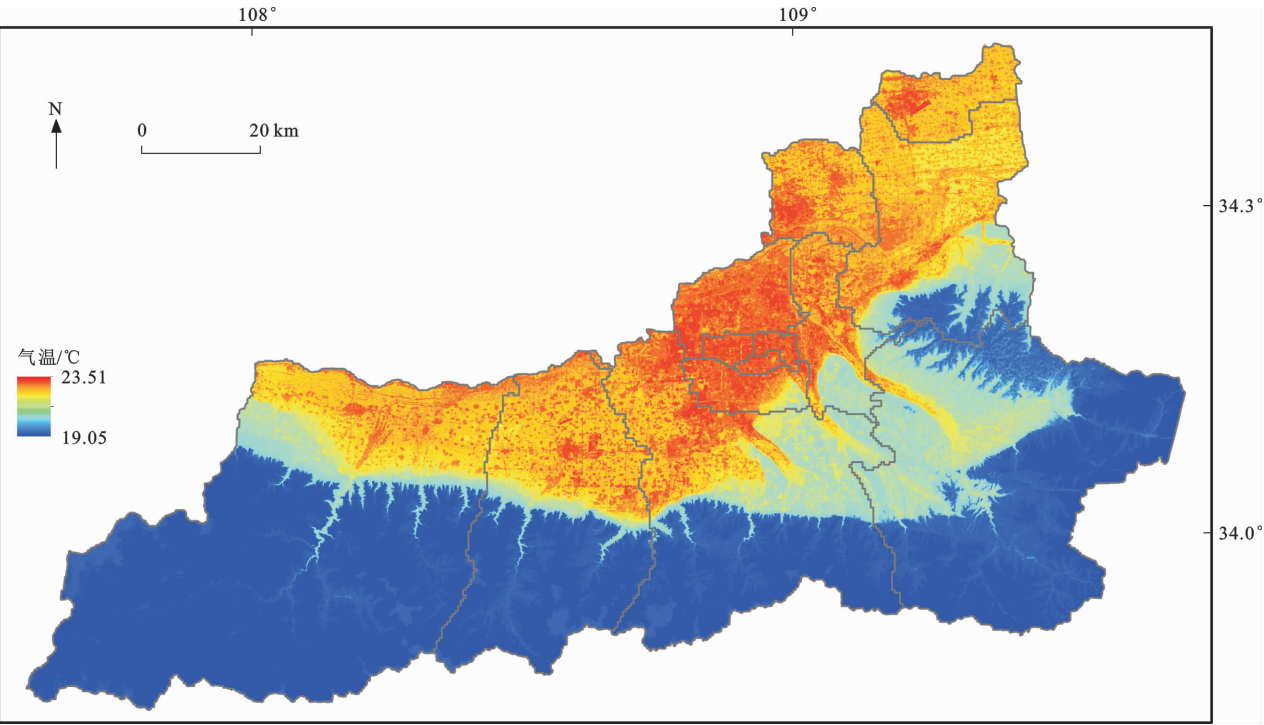


图 6 2016 年 5 月 16 日西安市近地表气温空间分布模拟图

Fig. 6 Spatial Distribution Simulation Map of the Near-surface Air Temperature in Xi'an City on May 16, 2016

区内灞河、浐河和渭河交汇处,水域面积广,植被覆盖度相对较高,因此,灞桥区平均气温较低。雁塔区大型景观公园较多,绿化率高,平均气温最低。莲湖区、新城区和碑林区属于老城区,建筑密集,人口密度大,平均气温相对较高。对于郊县区,在周至县、鄠邑区、长安区和蓝田县均观测到最低气温(19.05℃),明显低于高陵区和阎良区,与临潼区相差无异;其平均气温也明显低于高陵区和阎良区。秦岭位于西安市南部,东西向横跨周至县、鄠邑区、长安区和蓝田县,秦岭山区海拔高、植被覆盖度大,且以自然地表为主,因此,这 4 个区县平均气温低于高陵区、阎良区和临潼区。

综合图 6 和表 4 可知:西安市高温聚集在中心城区,最高气温为 23.51℃;郊县区以及植被较为浓密的地区气温低于中心城区;低温聚集在秦岭山区,最低气温为 19.05℃。西安市近地表气温呈现出中心城区较高、郊县区适中、山区较低的总体分布格局,且从中心城区到郊县区逐渐降低,山区气温普遍低于中心城区,呈现显著的城市热岛效应。

4 结 语

(1)相关性及其重要性综合分析表明,在地表温度、归一化植被指数、归一化建筑指数、改进的归一化水体指数和地表反照率等 5 个遥感因子以及高

程、坡向和坡度等 3 个地形因子中,高程对随机森林模型估算近地表气温的贡献度最大,其次是地表温度,坡向贡献度最低。

(2)7 种参数组合方案构建的随机森林模型均能够对训练集和验证集进行较好的拟合,训练集判定系数均高于 0.916,均方根误差均低于 0.467℃,验证集判定系数均高于 0.726,均方根误差均低于 0.840℃;训练集判定系数均高于验证集,均方根误差均低于验证集。P8 方案模型表现最优,其训练集判定系数为 0.934,均方根误差为 0.425℃,验证集判定系数为 0.795,均方根误差为 0.783℃,近地表气温估算精度判定系数为 0.792,均方根误差为 1.055℃,表明随机森林模型在研究区近地表气温估算中效果良好。

(3)2016 年 5 月 16 日,西安市平均气温为 21.98℃,13 个区县的最低气温平均值为 20.50℃,最高气温平均值为 23.49℃。中心城区的 3 个气温指标值均高于郊县区,其最低气温平均值、最高气温平均值及气温平均值分别高于郊县区 1.54℃、0.01℃和 1.76℃。水域和植被覆盖度高的地区气温相对较低。中心城区内,灞桥区气温低于其他辖区;郊县区内,周至县、鄠邑区、长安区和蓝田县低于高陵区、阎良区和临潼区。西安市近地表气温南低北高,空间差异明显,自中心城区至郊县区到南部山区逐

渐降低,呈现出显著的城市热岛效应。

(4)本文在近地表气温随机森林建模中综合考虑了地表温度、植被、水体、建筑物及地形等多种因素的影响,结合最优参数组合方案获取了精度较好的连续性空间近地表气温信息,对于开拓研究思路、提高估算精度、探索近地表气温遥感反演的新方法具有一定的参考意义,可为城市热岛效应研究提供新的思路。但是,影响近地表气温的因素众多,本文在模型设计时仅考虑了 5 个遥感因子和 3 个地形因子,未考虑太阳辐射、相对湿度、人口密度等自然因子及社会经济因子,后期将在模型中加入更多因子,深入探讨各因子对近地表气温估算的影响机理。此外,由于近地表气温实测气象站点数据有限,本文仅对单一时相数据进行了研究,下一步可在获取时间序列气温数据后开展近地表气温时空变化过程及机制研究,为城市生态环境保护和区域气候改善提供基础数据。

参考文献:

References:

- [1] PRIHODKO L, GOWARD S N. Estimation of Air Temperature from Remotely Sensed Surface Observations[J]. Remote Sensing of Environment, 1997, 60 (3):335-346.
- [2] 齐述华,王军邦,张庆员,等. 利用 MODIS 遥感影像获取近地层气温的方法研究[J]. 遥感学报, 2005, 9 (5):570-575.
QI Shu-hua, WANG Jun-bang, ZHANG Qing-yuan, et al. Study on the Estimation of Air Temperature from MODIS Data[J]. Journal of Remote Sensing, 2005, 9(5):570-575.
- [3] 祝善友,张桂欣,尹 球,等. 基于多源极轨气象卫星热红外数据的近地表气温反演研究[J]. 遥感技术与应用, 2009, 24(1):27-31.
ZHU Shan-you, ZHANG Gui-xin, YIN Qiu, et al. The Study on the Retrieval of the Air Temperature Based on Multi-sources Polar Orbit Meteorological Satellite Data[J]. Remote Sensing Technology and Application, 2009, 24(1):27-31.
- [4] 祝善友,张桂欣. 近地表气温遥感反演研究进展[J]. 地球科学进展, 2011, 26(7):724-730.
ZHU Shan-you, ZHANG Gui-xin. Progress in Near Surface Air Temperature Retrieved by Remote Sensing Technology[J]. Advances in Earth Science, 2011, 26(7):724-730.
- [5] CHANG Y, XIAO J F, LI X X, et al. Exploring Diurnal Thermal Variations in Urban Local Climate Zones

with ECOSTRESS Land Surface Temperature Data [J]. Remote Sensing of Environment, 2021, 263: 112544.

- [6] 冷 佩,廖前瑜,任 超,等. 近地表气温遥感反演方法综述[J]. 中国农业信息, 2019, 31(1):1-10.
LENG Pei, LIAO Qian-yu, REN Chao, et al. A Review of Methods for Estimating Near-surface Air Temperature from Remote Sensing Data[J]. China Agricultural Informatics, 2019, 31(1):1-10.
- [7] 孙从建,李 伟,陈 伟,等. 昆仑山提孜那甫河流域 2012~2016 年近地表气温时空分布特征[J]. 干旱区地理, 2019, 42(3):459-468.
SUN Cong-jian, LI Wei, CHEN Wei, et al. Spatiotemporal Distribution of Near-surface Temperature over the Tizinafu River Basin in the Kunlun Mountains from 2012 to 2016[J]. Arid Land Geography, 2019, 42 (3):459-468.
- [8] 游 婷,张 华,王海波,等. 夏季白天中国中东部不同类型云分布特征及其对近地表气温的影响[J]. 大气科学, 2020, 44(4):835-850.
YOU Ting, ZHANG Hua, WANG Hai-bo, et al. Distribution of Different Cloud Types and Their Effects on Near-surface Air Temperature During Summer Daytime in Central Eastern China[J]. Chinese Journal of Atmospheric Sciences, 2020, 44(4):835-850.
- [9] 李 军,游松财,黄敬峰. 中国 1961~2000 年月平均气温空间插值方法与空间分布[J]. 生态环境, 2006, 15(1):109-114.
LI Jun, YOU Song-cai, HUANG Jing-feng. Spatial Interpolation Method and Spatial Distribution Characteristics of Monthly Mean Temperature in China During 1961-2000[J]. Ecology and Environment, 2006, 15(1):109-114.
- [10] ALIDOOST F, STEIN A, SU Z B. Copula-based Interpolation Methods for Air Temperature Data Using Collocated Covariates[J]. Spatial Statistics, 2018, 28: 128-140.
- [11] FASSÒ A, SOMMER M, ROHDEN C V. Interpolation Uncertainty of Atmospheric Temperature Profiles[J]. Atmospheric Measurement Techniques, 2020, 13(12):6445-6458.
- [12] 韩秀珍,李三妹,窦芳丽. 气象卫星遥感地表温度推算近地表气温方法研究[J]. 气象学报, 2012, 70(5): 1107-1118.
HAN Xiu-zhen, LI San-mei, DOU Fang-li. Study of Obtaining High Resolution Near-surface Atmosphere Temperature by Using the Land Surface Temperature from Meteorological Satellite Data[J]. Acta Meteor-

- logica Sinica, 2012, 70(5): 1107-1118.
- [13] 周 佳, 赵亚鹏, 岳天祥, 等. 结合 HASM 和 GWR 方法的省级尺度近地表气温估算[J]. 地球信息科学学报, 2020, 22(10): 2098-2107.
ZHOU Jia, ZHAO Ya-peng, YUE Tian-xiang, et al. Near Surface Air Temperature Estimation by Combining HASM with GWR Model on a Provincial Scale[J]. Journal of Geo-information Science, 2020, 22(10): 2098-2107.
- [14] 徐伟燕, 孙 睿, 金志凤, 等. 基于 MODIS 数据的近地表气温估算[J]. 气象与环境科学, 2015, 38(1): 1-6.
XU Wei-yan, SUN Rui, JIN Zhi-feng, et al. Estimation of Near-surface Air Temperature Based on MODIS Data[J]. Meteorological and Environmental Sciences, 2015, 38(1): 1-6.
- [15] 陈命男. 上海城市地温的遥感反演及气温拟合研究[D]. 上海: 复旦大学, 2012.
CHEN Ming-nan. Research on Interpreting Land Surface Temperature and Fitting Air Temperature from Remote Sensing in Shanghai[D]. Shanghai: Fudan University, 2012.
- [16] LIN X H, ZHANG W, HUANG Y, et al. Empirical Estimation of Near-surface Air Temperature in China from MODIS LST Data by Considering Physiographic Features[J]. Remote Sensing, 2016, 8(8): 629-643.
- [17] MOHAMMADI C, FARAJZADEH M, RAHIMI Y G, et al. Comparison of Univariate and Multivariate Geographically Weighted Regression for Estimating Air Temperature over Iran[J]. Arabian Journal of Geosciences, 2018, 11(13): 360-376.
- [18] 徐永明, 覃志豪, 万洪秀. 热红外遥感反演近地层气温的研究进展[J]. 国土资源遥感, 2011(1): 9-14.
XU Yong-ming, QIN Zhi-hao, WAN Hong-xiu. Advances in the Study of Near-surface Air Temperature Retrieval from Thermal Infrared Remote Sensing[J]. Remote Sensing for Land and Resources, 2011(1): 9-14.
- [19] 徐永明, 覃志豪, 沈 艳. 基于 MODIS 数据的长江三角洲地区近地表气温遥感反演[J]. 农业工程学报, 2011, 27(9): 63-68.
XU Yong-ming, QIN Zhi-hao, SHEN Yan. Estimation of Near-surface Air Temperature from MODIS Data in the Yangtze River Delta[J]. Transactions of the CSAE, 2011, 27(9): 63-68.
- [20] PAPE R, LÖFFLE R. Modelling Spatio-temporal Near-surface Temperature Variation in High Mountain Landscapes[J]. Ecological Modelling, 2004, 178(3): 483-501.
- [21] YOO C, IM J, PARK S, et al. Estimation of Daily Maximum and Minimum Air Temperatures in Urban Landscapes Using MODIS Time Series Satellite Data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 137: 149-162.
- [22] 高 亮, 杜 鑫, 李强子, 等. 融合土地覆盖和土壤水分产品的近地表空气温度空间化方法[J]. 地球信息科学学报, 2020, 22(10): 2023-2037.
GAO Liang, DU Xin, LI Qiang-zi, et al. A Near-surface Air Temperature Spatialization Method Integrating Landuse and Soil Moisture Products[J]. Journal of Geo-information Science, 2020, 22(10): 2023-2037.
- [23] 邢立亭, 李 净, 焦文慧. 基于 MODIS 和随机森林的兰州市日最高气温和最低气温估算[J]. 干旱区研究, 2020, 37(3): 689-695.
XING Li-ting, LI Jing, JIAO Wen-hui. Estimation of Daily Maximum and Minimum Temperature of Lanzhou City Based on MODIS and Random Forest[J]. Arid Zone Research, 2020, 37(3): 689-695.
- [24] 陈孟奇. 基于 MODIS 数据的淮河流域近地表气温估算及年积温时空演变研究[D]. 南京: 南京师范大学, 2020.
CHEN Meng-qi. Study on Near-surface Air Temperature Estimation and Spital Temporal Evolution of the Accumulated Temperature in Huai River Basin Based on MODIS Data[D]. Nanjing: Nanjing Normal University, 2020.
- [25] 秦艳丽, 时 鹏, 何文虹, 等. 西安市城市化对景观格局及生态系统服务价值的影响[J]. 生态学报, 2020, 40(22): 8239-8250.
QIN Yan-li, SHI Peng, HE Wen-hong, et al. Influence of Urbanization on Landscape Pattern and Ecosystem Service Value in Xi'an City[J]. Acta Ecologica Sinica, 2020, 40(22): 8239-8250.
- [26] 孙 娴, 魏 娜, 郝 丽, 等. 近 60 年西安市极端气温事件变化特征分析[J]. 西北大学学报(自然科学版), 2014, 44(6): 997-1000.
SUN Xian, WEI Na, HAO Li, et al. Characteristics of Extreme Temperature Events in Xi'an for the Latest 60 Years[J]. Journal of Northwest University(Natural Science Edition), 2014, 44(6): 997-1000.
- [27] 金丽娜, 李雄飞. 2014~2017 年西安市城市热岛、冷岛精细化时空特征分析[J]. 沙漠与绿洲气象, 2021, 15(1): 97-102.
JIN Li-na, LI Xiong-fei. Spatial and Temporal Characteristics of Urban Heat Island and Cold Island in Xi'an City from 2014 to 2017[J]. Desert and Oasis

- Meteorology, 2021, 15(1): 97-102.
- [28] 聂敬娣, 张俊华, 黄 波. 城市热岛效应对人体健康影响研究综述[J]. 生态科学, 2021, 40(1): 200-208.
NIE Jing-di, ZHANG Jun-hua, HUANG Bo. A Review of the Human Health Consequences of Urban Heat Island Effect[J]. Ecological Science, 2021, 40(1): 200-208.
- [29] 白 琳, 徐永明, 何 苗, 等. 基于随机森林算法的近地表气温遥感反演研究[J]. 地球信息科学学报, 2017, 19(3): 390-397.
BAI Lin, XU Yong-ming, HE Miao, et al. Remote Sensing Inversion of Near-surface Air Temperature Based on Random Forest[J]. Journal of Geo-information Science, 2017, 19(3): 390-397.
- [30] 胡德勇, 乔 琨, 王兴玲, 等. 单窗算法结合 Landsat 8 热红外数据反演地表温度[J]. 遥感学报, 2015, 19(6): 964-976.
HU De-yong, QIAO Kun, WANG Xing-ling, et al. Land Surface Temperature Retrieval from Landsat 8 Thermal Infrared Data Using Mono-window Algorithm[J]. Journal of Remote Sensing, 2015, 19(6): 964-976.
- [31] 胡德勇, 乔 琨, 王兴玲, 等. 利用单窗算法反演 Landsat 8 TIRS 数据地表温度[J]. 武汉大学学报(信息科学版), 2017, 42(7): 869-876.
HU De-yong, QIAO Kun, WANG Xing-ling, et al. Comparison of Three Single-window Algorithms for Retrieving Land-surface Temperature with Landsat 8 TIRS Data[J]. Geomatics and Information Science of Wuhan University, 2017, 42(7): 869-876.
- [32] ROUSE J W, HAAS R H, SCHELL J A, et al. Monitoring Vegetation Systems in the Great Plains with ERTS[C]//FREDEN S C, MERCANTI E P, BECKER M A. The Third Earth Resources Technology Satellite-1 Symposium. Washington DC: NASA, 1974: 309-317.
- [33] 徐涵秋. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究[J]. 遥感学报, 2005, 9(5): 589-595.
XU Han-qiu. A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index(MNDWI)[J]. Journal of Remote Sensing, 2005, 9(5): 589-595.
- [34] 查 勇, 倪绍祥, 杨 山. 一种利用 TM 图像自动提取城镇用地信息的有效方法[J]. 遥感学报, 2003, 7(1): 37-40.
ZHA Yong, NI Shao-xiang, YANG Shan. An Effective Approach to Automatically Extract Urban Land-use from TM Imagery [J]. Journal of Remote Sensing, 2003, 7(1): 37-40.
- [35] LIANG S L. Narrowband to Broadband Conversions of Land Surface Albedo I Algorithms [J]. Remote Sensing of Environment, 2001, 76(2): 213-238.
- [36] BREIMAN L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [37] 李永丽, 王 浩, 金喜子. 基于随机森林优化的自组织神经网络算法[J]. 吉林大学学报(理学版), 2021, 59(2): 351-358.
LI Yong-li, WANG Hao, JIN Xi-zi. Self-organizing Neural Network Algorithm Based on Random Forest Optimization[J]. Journal of Jilin University (Science Edition), 2021, 59(2): 351-358.
- [38] VLADIMIR S, ANDY L, CHRISTOPHER T, et al. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling[J]. Journal of Chemical Information and Computer Sciences, 2003, 43(6): 1947-1958.
- [39] MIN M, LI J, WANG F, et al. Retrieval of Cloud Top Properties from Advanced Geostationary Satellite Imager Measurements Based on Machine Learning Algorithms[J]. Remote Sensing of Environment, 2020, 239: 111616.
- [40] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
FANG Kuang-nan, WU Jian-bin, ZHU Jian-ping, et al. A Review of Technologies on Random Forests[J]. Statistics and Information Forum, 2011, 26(3): 32-38.
- [41] CHOU J S, PHAM A D. Nature-inspired Metaheuristic Optimization in Least Squares Support Vector Regression for Obtaining Bridge Scour Information[J]. Information Sciences, 2017, 399: 64-80.
- [42] LIU Y, BI J W, FAN Z P. Multi-class Sentiment Classification: The Experimental Comparisons of Feature Selection and Machine Learning Algorithms[J]. Expert Systems with Applications, 2017, 80: 323-339.
- [43] KOHAVI R. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection [C]// IJCAI. Proceedings of the 14th International Joint Conference on Artificial Intelligence. Montreal: IJCAI, 1995: 1137-1143.