

# 基于随机森林模型的太阳辐射中长期变化分析

贾兴斌<sup>1,2</sup>, 汪国菊<sup>1,2</sup>, 王仁政<sup>3</sup>, 宫响<sup>1,2</sup>

(1. 青岛科技大学数理学院, 青岛 266061; 2. 青岛市人工智能海洋技术创新中心, 青岛 266061;

3. 中国海洋大学环境科学与工程学院, 青岛 266100)

**摘要:** 该文基于多源辐射观测资料,采用随机森林(RF)算法、季节差分自回归移动平均(SARIMA)模型及特征重要性等方法,对山东省济南市太阳辐射长期变化趋势和影响因素进行综合分析。结果显示:RF模型拟合月太阳辐射效果较好,决定系数和平均绝对百分比误差分别为0.92和9%,优于SARIMA模型;济南市及周边地区月太阳辐射1980—2020年经历“变暗”到“变亮”的过程,空间呈现西北高东南低的特点;最高温度和日照时数是影响太阳辐射月变化拟合准确度的主要因素,降雨量是导致月太阳辐射总量突变的重要原因,大气污染物中SO<sub>2</sub>和O<sub>3</sub>与太阳辐射的相关性最大。

**关键词:** 太阳辐射; 随机森林; 特征选择; SARIMA模型; 拟合分析

中图分类号: TK519

文献标志码: A

## 0 引言

太阳能具有环保、可持续等优点,是应对能源短缺、气候变化的重要选择<sup>[1]</sup>。分析太阳辐射是模拟和设计各种太阳能系统的前提<sup>[2-6]</sup>。太阳辐射变化与温度、蒸发、空气污染等密切相关<sup>[7-10]</sup>,对人类生存环境和生态系统有重要影响<sup>[11]</sup>。山东省太阳能资源丰富,但受限于观测站数和传统产业群<sup>[12]</sup>,太阳辐射的理论分析仍无法有效指导太阳能产业的发展。

国内外学者主要使用数值模型、经验模型和人工智能模型等预测地面太阳辐射<sup>[13-14]</sup>。数值模型计算复杂,经验模型仅适用于局部,相比之下,人工智能模型具有精度高、效率快等优点<sup>[15-21]</sup>。其中随机森林(random forest, RF)模型模拟效果较好<sup>[22-23]</sup>,该模型不仅可预测太阳辐射,还能深入挖掘预测因子与太阳辐射的交互关系,具有较高的机理分析价值<sup>[24]</sup>。

本文基于RF模型,利用多种气象要素拟合分析月太阳辐射总量,并对比传统季节差分自回归移动平均(seasonal auto regression integrated moving average, SARIMA)模型,RF模型的准确性。其次,结合RF模型的特征重要性和热力图分析太阳辐射与济南市气候和环境的相互影响,并利用克里金插值方法,分析济南及周边区域太阳辐射的时空变化,以期对太阳能产业的发展提供重要理论参考。

## 1 数据与方法

### 1.1 数据准备与预处理

本文所采用数据包括下载自中国气象数据网的地面气

象资料和辐射资料月值数据(<http://data.cma.cn/data/>),来自欧洲中期天气预报中心(European centre for medium-range weather forecasting, ECMWF)第5代大气再分析(ERA5)月平均直接辐射数据(<https://cds.climate.copernicus.eu/>),以及下载自哥白尼大气监测服务中心的全球再分析(ECMWF Atmospheric Composition Reanalysis 4, EAC4)的环境月平均数据(<https://ads.atmosphere.copernicus.eu/>)。其中,气象数据和辐射数据时间为1980年1月—2017年3月,气候数据集包含风速(极大风速、平均2 min 风速)、气温(平均气温、最低气温、最高气温、平均最低气温、平均最高气温)、气压(平均气压、平均水汽压)、平均相对湿度、降水量(20时—次日20时降水量、最大日降水量、日降水量≥0.1 mm 日数)、日照时数和总云量共15种类型。环境数据选用时间范围为2003年1月—2017年3月,由CO、NO<sub>2</sub>、SO<sub>2</sub>、O<sub>3</sub>、PM<sub>2.5</sub>、PM<sub>10</sub>组成;直接辐射数据空间分辨率为0.5°×0.5°,选择时间为1980年1月—2020年12月。

本文选择3 $\sigma$ 法则对数据集进行异常值判断,此外由于气候和环境数据集中异常值和缺失值在总数中占比较少(小于1%),因此采用均值代替的方法处理异常值和缺失值。数据处理后选择气候数据和辐射数据中1980年1月—2011年7月的381条数据(85%)训练RF模型和SARIMA模型,选择2018年8月—2017年3月的66条数据(15%)用于模型检测。

### 1.2 分析方法

#### 1.2.1 随机森林模型

RF模型是机器学习方法之一<sup>[22]</sup>,其本质上是基于决策

收稿日期: 2023-01-06

基金项目: 国家自然科学基金-山东省联合基金(U1906215); 青岛科技大学研究生自主创新研究专项(S2022KY029)

通信作者: 宫响(1977—),女,博士、副教授,主要从事海洋大气大数据分析方面的研究。gongxiang@qust.edu.cn

树的集成学习模型,结构如图1所示。从图1中可看出,同一个样本在不同树中所归属的叶子节点也不尽相同,甚至连类别也可能不同。但这也充分体现了“Bagging”集成模型的优点,即通过“平均”来提高模型的泛化能力。

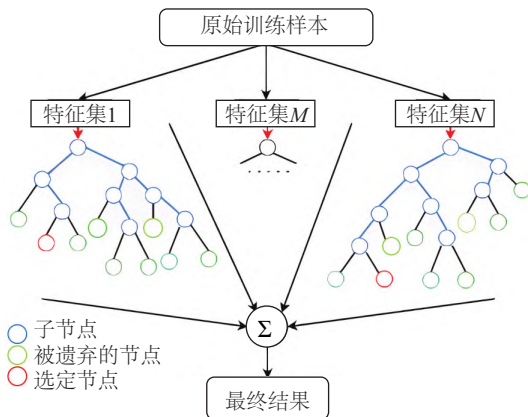


图1 随机森林结构示意图

Fig. 1 Structure of random forest

RF的每个决策树在每次划分节点的过程中,只会从已有的特征中再随机选择部分特征参与节点划分,通过这一过程,决策树在节点划分时就会选择同样的特征,使得最终得到的决策树之间具有较强的关联性,即每棵树都类似。

本文构建RF分析模型,通过式(1)计算因子得分,结合相关系数热力图分析其他气象因素和环境因素与月总太阳辐射的相互影响。

$$S_{imp} = \frac{N}{N} \times \left( G_t - \frac{N_{il}}{N_t} G_{il} - \frac{N_{ir}}{N_t} G_{ir} \right) \quad (1)$$

式中:  $N$ ——样本数量;  $N_t$ —— $t$ 节点的样本数量;  $G_t$ —— $t$ 节点的纯度;  $N_{il}$ —— $t$ 节点左子树中的样本数量;  $G_{il}$ —— $t$ 节点左子树纯度;  $N_{ir}$ —— $t$ 节点右子树中的样本数量;  $G_{ir}$ —— $t$ 节点右子树的纯度。

### 1.2.2 季节差分自回归移动平均预测模型

SARIMA模型<sup>[25-26]</sup>常被用于拟合季节性周期变化(包括季度、月度、周等)的时间序列,该模型在ARIMA模型中引入了季节性项( $P, D, Q$ ),可表示为ARIMA( $p, d, q$ ) $\times$ ( $P, D, Q$ ) $m$ ,其中,ARIMA( $p, d, q$ )可表示为:

$$\Phi(B)(1-B)^d Y_t = \Theta(B) a_t \quad (2)$$

式中:  $B$ ——延迟算子;  $\Phi(B)$ ——ARIMA( $p, d, q$ )模型的自回归系数多项式;  $\Theta(B)$ ——ARIMA( $p, d, q$ )模型的移动平滑系数多项式;  $Y_t$ —— $T$ 时刻模型预测值;  $a_t$ ——零均值白噪声序列。

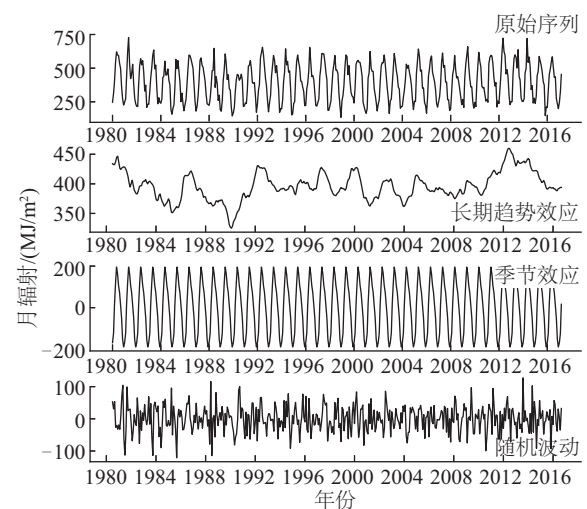
SARIMA模型中 $p, d, q$ 分别代表原始序列中自回归项、差分项和移动平均项, $P, D, Q$ 分别代表季节序列中的自回归项、差分项和移动平均项, $m$ 是每年的观测数量。季节性部分与非季节性相似,但考虑了季节性时间回溯。

## 2 结果与讨论

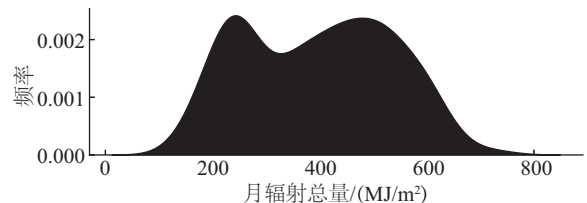
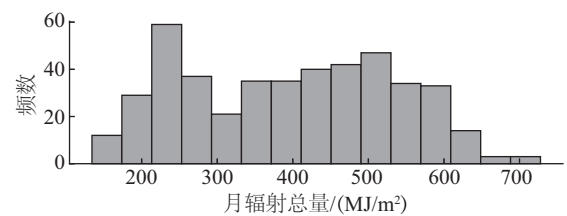
### 2.1 太阳总辐射的季节变化

首先对处理好的观测值序列,绘制时间序列及季节-趋势分解图,如图2a所示。从图中可看到时间序列有比较明显的季节性趋势,即每一年数据都呈现先上升再下降的形状,同时图2a中显示时间序列存在复杂的季节效应(seasonal)、长期趋势效应(trend)和随机波动(resid)之间的交互影响关系,且其时间序列周期(seasonal)规则为12个月。

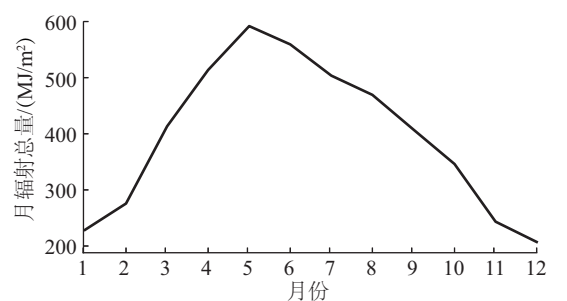
图2b是原始时间序列的分布图,从图中可明显得到,从1980年起月总太阳辐射总体呈现双峰分布,峰值区间在200~300和400~600 MJ/m<sup>2</sup>范围内,主要分布时间分别为每年



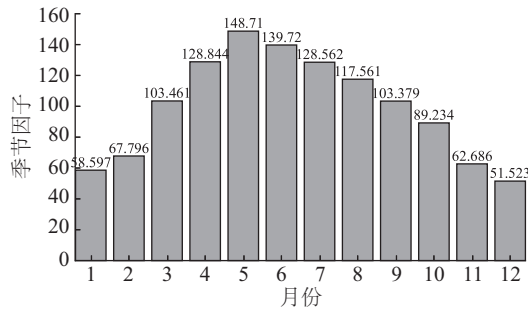
a. 原始序列季节-趋势分解



b. 序列频数-频率分布



c. 月平均太阳辐射总量



d. 月太阳辐射季节因子

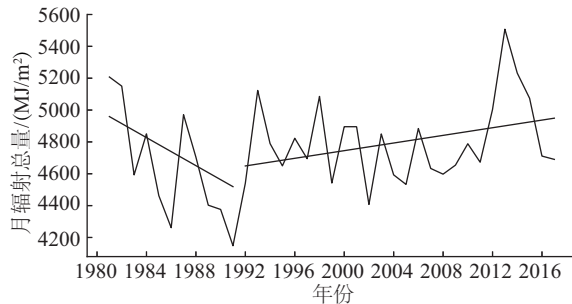
图2 时间序列信息

Fig. 2 Time series information

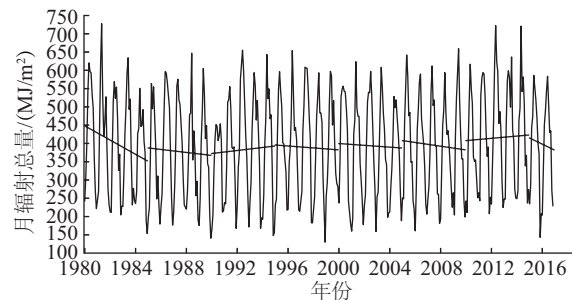
的11—2月份和每年的4—8月份。图2c为月太阳辐射平均值分布图。图中显示春季占一年辐射总量的25%，夏季占一年辐射总量的35%，秋季占辐射总量的26%，冬季占辐射总量的14%。图2d对原始序列做季节因子分析，从季节因子看，每年12月最低，5月最高。

### 2.2 太阳辐射年际变化及气象影响因素分析

太阳辐射对气候和环境有重要影响，近几十年，全球绝大部分地区的地表太阳辐射总量都经历了一个从减少到增加的过程，也就是所谓的地球“变暗”到“变亮”<sup>[9]</sup>。济南地区太阳辐射变化与全球范围内太阳辐射“变暗”及“变亮”的变化是一致的，如图3所示，1990年之前，太阳辐射呈明显的下降趋势；1990年之后，太阳辐射虽然具有波动性，但总体呈上升趋势。气候变化是引起这种趋势变化的重要因素<sup>[27]</sup>。



a. 长期趋势



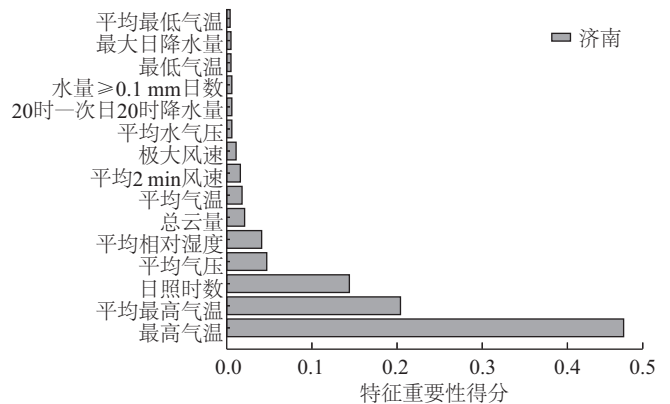
b. 每5年趋势

图3 1980—2016年济南市太阳总辐射时间序列趋势

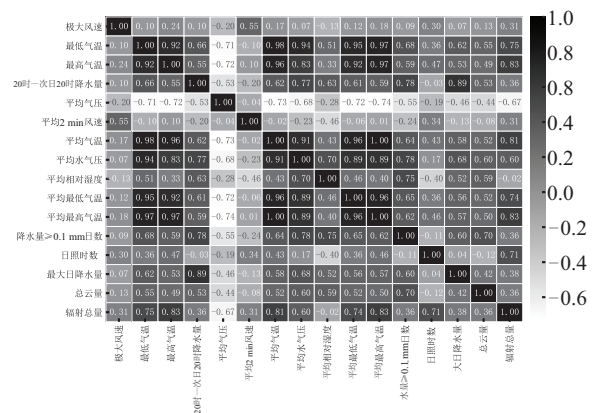
Fig. 3 1980-2016 time series trend of total solar radiation in Ji'nan

Bristow等<sup>[28]</sup>研究发现太阳辐射与空气温度的日范围有关，并建立一个简单的通过温度预测日太阳辐射变化的模型；Samani<sup>[29]</sup>同样使用最高和最低温度来估计太阳辐射。

分析济南地区各气候因子与太阳辐射的交互影响关系，利用式(1)计算各气候因子对太阳辐射的特征重要性，结果如图4a所示。在所有15种气候因子中，最高气温与太阳辐射的交互影响最大，其次为平均最高气温和日照时数。在温度类(平均气温、最低气温、最高气温、平均最低气温、平均最高气温)、风速类(极大风速、平均2 min风速)、气压类(平均气压、平均水气压)、降水类(20时一次日20时降水量、最大日降水量、日降水量≥0.1 mm日数，平均相对湿度)、其他类(日照时数、总云量)5类气候因子中，温度类与太阳辐射的交互影响，具有决定性作用。图4b的相关系数热力图中可看到最高气温、平均最高气温和太阳辐射的相关系数都达到0.83，且分析观测值数据，1990年前济南地区最高气温平均值为26.03℃，1990年后最高气温平均值上升至26.93℃，最高气温平均值上升0.9℃，变化趋势与太阳辐射类似，济南地区气温的长期变化可用太阳辐射的变化来解释，同时通过对温度的长期趋势的分析可获得太阳辐射的长期变化。



a. 气候因子特征重要性



b. 气候因子相关系数热力图

图4 气候因子特征重要性及相关系数热力图

Fig. 4 Importance of climate factor characteristics and correlation coefficient heat map

日照时数<sup>[30]</sup>作为太阳总辐射量的计算指标之一，与太阳

辐射总量具有强正相关性,日照时间越长,获得太阳辐射越强。同时本文通过图 4b 发现,太阳辐射与大气压有明显的负相关,1990 年前济南地区平均气压为 1010.21 hPa,1990 年之后平均气压为 1000.76 hPa,下降了 1%,而平均太阳辐射上升了 0.3%。分析其原因,可能大气辐射影响大气环流后,进而影响到大气辐射在大气中的吸收、反射作用等。

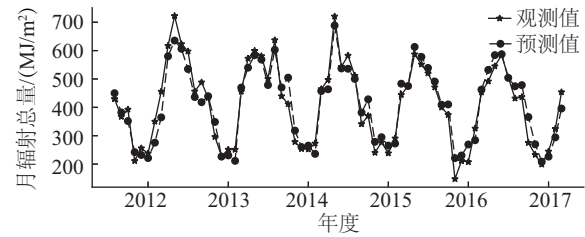
降水类气候因子虽然总体对太阳辐射影响较小,但会导致太阳辐射产生较大波动。1980—2017 年,太阳辐射春季突变点为 2001 年 1 月,夏季突变点为 2014 年 5 月,秋季突变点为 1985 年 9 月,冬季突变点为 2015 年 11 月,分析其突变原因主要由降雨造成,当天气不稳定,降雨比较多的月份,过多的云层会影响到日照时数,进而使太阳辐射波动剧烈<sup>[31]</sup>;相反降雨少则会引起太阳辐射的上升。例如 1985 年济南年降雨量达到 708 mm,且主要集中在 9 月份,这也使得 9 月份日照时数由 8 月份的 224.2 h 迅速减少到 136.0 h,导致 9 月份太阳辐射产生突变;同样 2014 年济南年降雨量由 736 mm 减少到 441 mm,特别是 5 月份降雨天数大幅减小,日照时数由 4 月份的 193.8 h 增加到 302.3 h,济南太阳辐射增加明显,产生突变。

总云量与太阳辐射密切相关,它们之间有着复杂的关系<sup>[9,27]</sup>。云量用来衡量太阳辐射的大小,而随着云量的增加,太阳辐射也会减少<sup>[27]</sup>。图 4a 和图 4b 中总云量和月辐射总量有明显的相关关系。一般来说,云量的增加会导致太阳辐射的减少。因为云是大气中的一种热稳定因子,它能吸收、延迟和减弱太阳辐射,从而降低地表太阳辐射的强度使得大气总能量也减少<sup>[32]</sup>。云量的减少也会影响太阳辐射的大小。由于云的减少,大气中的热稳定因子减少,太阳辐射就不会被减弱,大气总能量就会增加。济南地区 1990 年前,年平均总云量为 0.5,1990 年之后年平均总云量是 0.475,这一变化规律与太阳辐射变化相似。这种交互影响关系对于许多气象现象,如气温、降水等都有重要影响<sup>[33]</sup>。

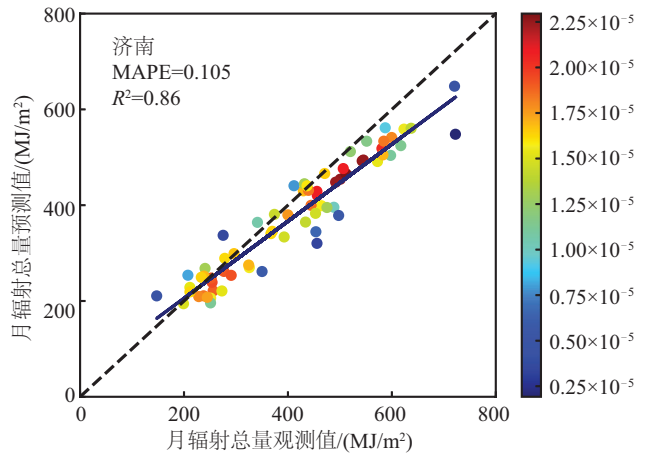
### 2.3 太阳总辐射估算结果

本文使用 RF 算法,首先利用平均气温、风速、平均气压、相对湿度等 15 个气象因素估算月总太阳辐射量。同样选择 1980 年 1 月—2011 年 7 月的 381 条数据用于训练,2018 年 8 月—2017 年 3 月的 66 条数据用于检测。模型通过网格化参数寻优,选择随机森林中决策树的数量( $n_{estimator}$ )为 500,决策树的最大深度( $max\_depth$ )为 20,寻找最佳分裂时要考虑的特征数量( $max\_features$ )为 5 的训练模型。模型测试效果如图 5 所示。

图 5a 为测试集拟合情况,图 5b 为拟合值与真实值的密度图。从图 5a 看出在 2014 年之前预测值较真实值波动较大,实线和虚线吻合程度相对不理想外,如 2012 年 5—9 月份、2013 年 3—5 月份等拟合值较真实值偏低,其他时间模型都起到了较好的拟合效果,特别是在 2014—2017 年。图 5b 为中拟合值和真实值都贴近 1:1 直线附近(图中虚线),且拟



a. 随机森林模型测试集拟合图



b. 随机森林模型测试集密度图

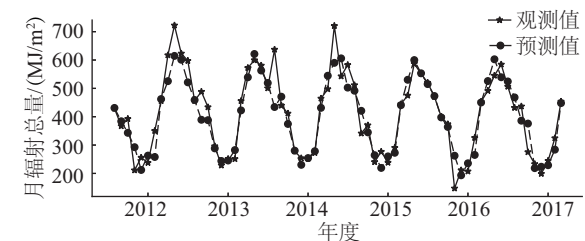
图 5 随机森林模型测试

Fig. 5 Random forest model testing

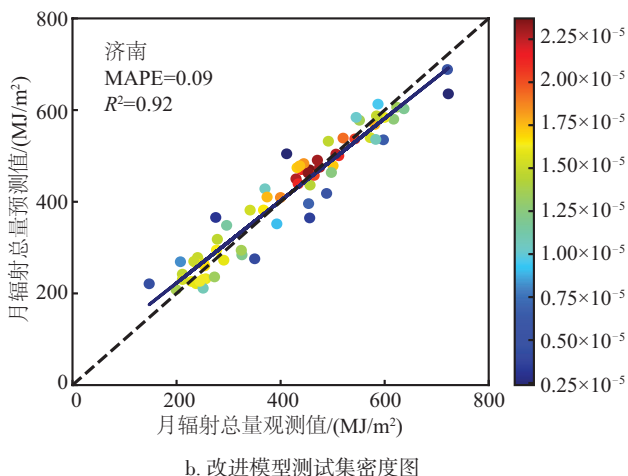
合值的回归直线(图中实线)接近 1:1 直线。图 5b 中平均绝对百分比误差(mean absolute percentage error, MAPE)约为 0.1,表明模型拟合结果和真实数据较为接近; $R^2$  为 0.86,反映出 RF 模型具有较高的准确性,这表明该模型可用于济南市月总太阳辐射的拟合分析。

为进一步验证太阳辐射与气象因素的交互影响及提高拟合精度,考虑特征重要性大于 0.05 的变量作为随机森林模型的拟合因子,再次训练测试 RF 模型。图 4a 中 15 个因子中最高气温、平均最高气温、日照时数、平均气压的特征重要值均大于 0.05,故取这 4 个变量作为主要拟合因子,同时虽然降水类因子重要性在 0.05 以下,但考虑其对辐射突变情况的反映,也选取日降水量 $\geq 0.1$  mm 日数和平均相对湿度为 RF 模型的输入变量。

经特征选择改进后的 RF 模型测试结果如图 6 所示,改进输入后 RF 模型拟合效果有很大提升,图 6a 中实线和虚线吻合程度高,图 6b 显示:改进模型的  $R^2$  达到 0.92,而 MAPE 则下降到了 0.09,模型拟合精度有较大程度提升。



a. 改进模型测试集拟合图



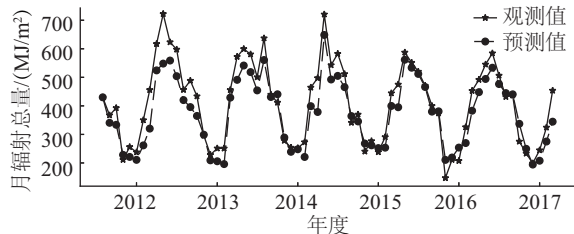
b. 改进模型测试集密度图

图6 改进随机森林模型测试

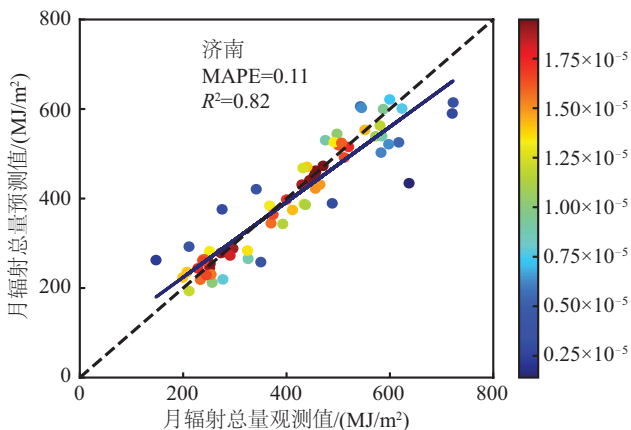
Fig. 6 Improved random forest model testing

### 2.4 SARIMA 模型对比

本文选择 AIC 准则确定模型阶数<sup>[25]</sup>, 取  $p \in [0, 40]$ ,  $q \in [0, 40]$ , 选取使 AIC 达到最小的  $p, q$  阶数, 最终确定模型为 SARIMA(1,0,0)(1,0,1)<sup>[12]</sup>。再对模型进行参数和残差检验, 判断 SARIMA(1,0,0)(1,0,1)<sup>[12]</sup>通过检验后, 将模型向前预测 66 期, 即预测 2011 年 8 月—2017 年 3 月, 并将预测结果与测试集进行对比, 结果如图 7 所示。图 7a 显示在个别月份, 如 2012 年 5—7 月份、2013 年 8 月份、2014 年 5 月份等时间预测值偏低, 其他时段模型也有较好的预测效果。图 7b 显示模型 MAPE 为 11%,  $R^2$  为 0.82, 模型也可用于



a. SARIMA 模型测试集拟合图



b. SARIMA 模型测试集密度图

图7 SARIMA 模型测试

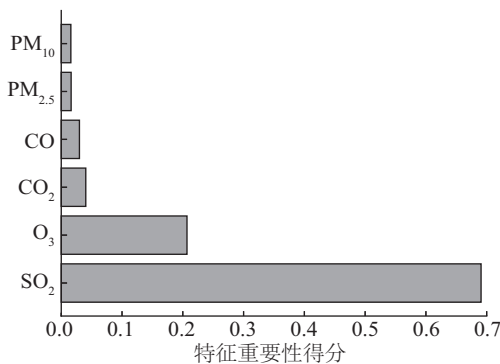
Fig. 7 SARIMA model testing

对太阳辐射的预测。对比随机森林模型。SARIMA 模型 MAPE 略有降低,  $R^2$  下降了 4 个百分点, 而对比改进后的随机森林模型,  $R^2$  则下降了 10 个百分点, 对比结果进一步验证了改进随机模型良好的拟合效果。

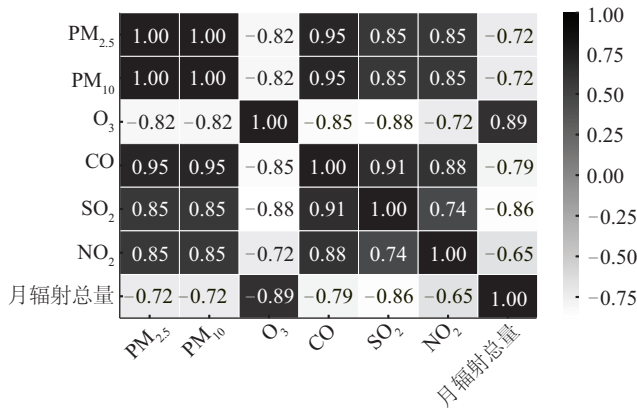
气候的不稳定性也会造成模型拟合精度的下降, 这可能也是随机森林模型在 2014 年 6 月前拟合精度不够的主要原因, 而 SARIMA 模型未考虑气候变量协助预测, 提高了模型的稳定性, 但同样对突变点出的预测 SARIMA 模型也未达到很好的效果, 如 2013 年 8 月份、2014 年 5 月份等。

### 2.5 大气污染物对太阳辐射的影响

地表太阳辐射不仅受到气候因素的影响, 同样受到人类活动和大气污染的影响<sup>[34]</sup>。薛德强<sup>[35]</sup>认为济南市 1961—1990 年地表太阳辐射量的减少, 主要是由大气污染物增多导致的; 王建源等<sup>[36]</sup>发现受环境影响山东省 2001—2007 年总辐射较前 30 年平均减少 72.3 MJ/m<sup>2</sup>。图 8a 所示济南市 6 种大气污染因子中, 4 种气态污染物对太阳辐射的影响均要高于 PM<sub>2.5</sub> 和 PM<sub>10</sub> 两种颗粒污染物, SO<sub>2</sub> (特征得分 0.69) 和 O<sub>3</sub> (特征得分 0.21) 是影响济南市太阳辐射的最主要的污染物影响因子。图 8b 显示除 O<sub>3</sub> 外, 其他 5 种污染物因子均与太阳辐射呈现负相关, 即这些污染物因子大量排放都会导致太阳辐射的下降, 在济南地区尤其以 SO<sub>2</sub> 为主。SO<sub>2</sub> 发生光化学反应时要吸收大量的太阳辐射, 而济南地区作为主要的工业城市, 大部分工业区扩散条件差, 排放的 SO<sub>2</sub> 排放量占全市大气污染重点污染源的 90%<sup>[37]</sup>。从 1980 年改革开放开始, 在很长的一段时间内 SO<sub>2</sub> 的过量排放是导致济南市太阳辐射总量下降的重要原因之一。自 2013 年开始, 山东省加大了对环境治理的力度, 济南市空气中 SO<sub>2</sub> 含量逐年下降<sup>[37]</sup>, 这也是济南市太阳辐射总量稳中有升的重要原因。与 SO<sub>2</sub> 和 PM<sub>2.5</sub> 等颗粒污染物不同, 地表 O<sub>3</sub> 是一个很活跃的气体, 主要由氮氧化物和挥发性有机物在高温、日照充足的环境下通过光化学反应生成, 在太阳辐射充足的月份, 地面臭氧浓度会上升, 过量的 O<sub>3</sub> 则会影响居民健康和安全<sup>[38]</sup>。因此如欲发展新能源产业, 进一步坚持和完善各项大气污染防治措施, 协调太阳能产业发展和环境综合治理并行才能更好的进行工业、能源业转型和升级。



a. 环境污染因子特征重要性



b. 环境污染因子相关系数热力图

图8 环境污染因子特征重要性及相关系数热力图

Fig. 8 Heat map of importance of environmental pollutant factor characteristics and correlation coefficients

### 2.6 太阳直接辐射的时空分布分析

限于中国气象网辐射数据,山东省只有福山、烟台、济南、济宁、莒县 5 个站点,数据量少,空间分布稀疏,对太阳辐射进行时空分析较为困难。本文选择欧洲中期天气预报中心的月均太阳直接辐射数据,选择济南市及济南周边德州陵

城区、齐河县、肥城县、商河县、泰安岱岳区、邹平县、惠民县、济南莱芜区 9 个地点,通过克里金插值法<sup>[39]</sup>获得济南市太阳辐射年值分布如图 9 所示。

济南市及周边区域 1980 年、2000 年、2010 年、2015 年、2020 年这 5 年的太阳直接辐射在 3600 MJ/m<sup>2</sup> 上下波动,1990 年太阳直接辐射均值只有 3400 MJ/m<sup>2</sup>,与济南市长期太阳辐射总量的先“变暗”后“变亮”的趋势变化相符合。

从空间分布来看,济南主要呈现北高南低、西高东低的情况,这与王建源等<sup>[36]</sup>对山东省太阳总辐射年变化表现北部多、南部少的结论相符,其中济南市太阳直接辐射的最高值出现在 2015 年的商河县,最低值出现在 1990 年的莱芜区,同年度济南各市区有较明显差异。分析其原因,与山东省地形有极大的关系。山东省地势总体上西南高东北低,中南部地区为山地丘陵,其他地区多为平原地形。济南恰处于山东省山地地带,地形大体可分为两类:北部、中部(商河、天桥等)以平原为主,南部(平阴、莱芜等)以丘陵山地为主<sup>[40-41]</sup>。莱芜区处于山谷缓冲地带,根据沈志宝等<sup>[41]</sup>的研究,山谷地带会削弱太阳辐射,因此 40 年来,济南地区太阳辐射的最低值都在莱芜区,不适合发展大规模的太阳能产业,商河县,地势开阔,太阳辐射充足、稳定,有利于发展太阳能产业。

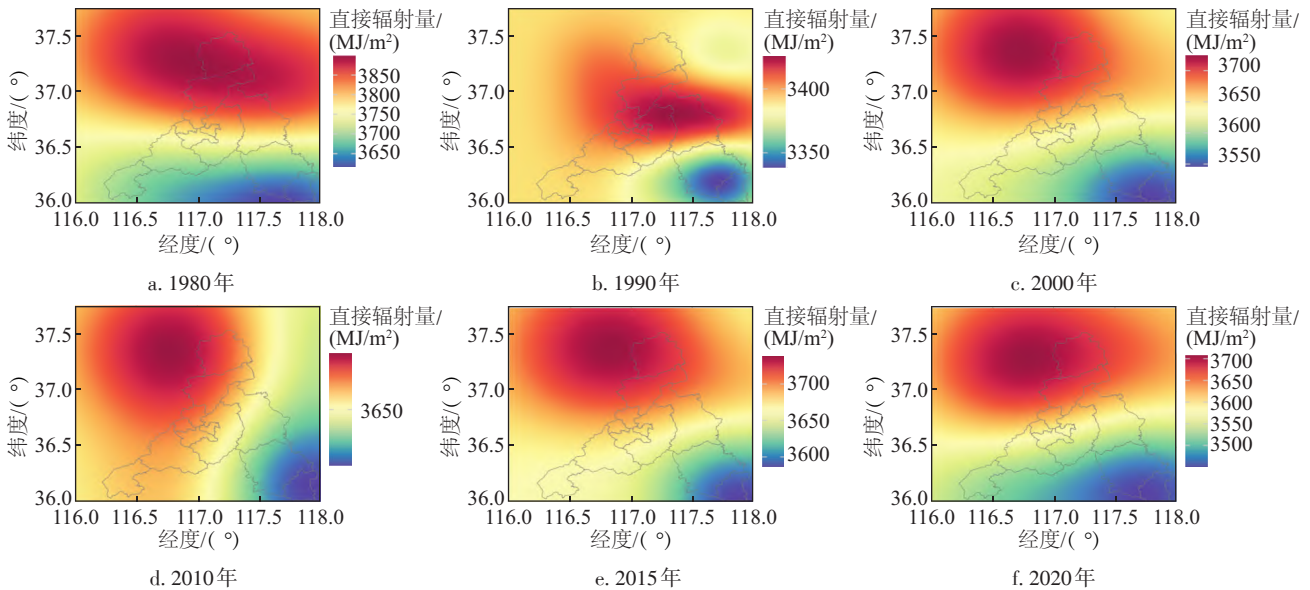


图9 济南市直接辐射时空分布图

Fig. 9 Spatial and temporal distribution of direct radiation in Ji'nan

## 3 结论

本文使用多源的太阳辐射、直接辐射和地面气候资料数据,对晋冀鲁地区重要工业城市——济南的太阳辐射做全面系统分析。本文首先建立 RF 预测模型,依据多种气候预报因子预测月太阳辐射总值,实现了对太阳辐射的精准拟合分析,模型决定系数 R<sup>2</sup> 达到 0.86;其次通过特征因子重要性指标和相关系数热力图,全面分析太阳辐射与气候因子和环境

要素之间的相互影响,揭示在济南站点,温度类气候因子对太阳辐射的决定性影响作用及降水导致太阳辐射出现不规则波动等,根据分析结果改进随机森林分析模型输入因子,改进随机森林模型的 R<sup>2</sup> 达到 0.9 以上,对比传统的 SARIMA 预测模型 MAPE 降低了 2 个百分点, R<sup>2</sup> 提高了 10 个百分点,模型更加稳定和可靠;结合济南市实际污染情况从理化性质系统分析太阳辐射对 SO<sub>2</sub> 和 O<sub>3</sub> 的交互影响。最后本文利用克里金插值法,从太阳直接辐射角度,得出济南市太阳

辐射西高东低、北高南低的结论。

本文研究结果对指导济南市太阳能产业发展具有重要理论价值:

1)根据结果反映的太阳辐射量的季节变化特征,针对济南地区不同季节,可针对性地设计太阳能发电系统,更好地满足居民对电力的需求。从而为太阳能发电的投资可行性提供重要的参考依据。

2)太阳辐射数据反映的与环境条件和气候条件的交互关系,可更好地了解太阳能发电系统的设计与运行状态,以及太阳能资源的利用效率。更好协调环境治理与太阳能产业发展。

3)依据插值结果反映的太阳辐射量的时空分布特征,可为太阳能产业的发展提供重要的参考依据。根据插值结果,政府可集中在太阳能资源密集的高河地区投资建设太阳能相关产业,以达到节能减排的目标。

本文也存在一些不足之处,如未深入分析风速类气候因子与太阳辐射的关系、未对更广阔区域的太阳辐射作分析,无足够的污染数据进一步提升随机森林预报模型的精度等,这些问题有待后续做进一步研究。

#### [参考文献]

- [1] 闫云飞,张智恩,张力,等.太阳能利用技术及其应用[J].太阳能学报,2012,33(S1):47-56.  
YAN Y F, ZHANG Z E, ZHANG L, et al. Application and utilization technology of solar energy[J]. Acta energiae solaris sinica, 2012, 33(S1): 47-56.
- [2] TAHIR Z R, ASIM M. Surface measured solar radiation data and solar energy resource assessment of Pakistan: a review[J]. Renewable and sustainable energy reviews, 2018, 81: 2839-2861.
- [3] COSKUN C, OKTAY Z, DINCER I. Estimation of monthly solar radiation distribution for solar energy system analysis[J]. Energy, 2011, 36(2): 1319-1323.
- [4] 毛前军,谢鸣,帅永,等.太阳辐射强度对太阳能腔式吸热器热流密度的影响[J].太阳能学报,2013,34(10):1818-1822.  
MAO Q J, XIE M, SHUAI Y, et al. Effect of incident solar irradiation on radiation flux distribution of a solar cavity receiver[J]. Acta energiae solaris sinica, 2013, 34(10): 1818-1822.
- [5] MYERS D R. Solar radiation modeling and measurements for renewable energy applications: data and model quality [J]. Energy, 2005, 30(9): 1517-1531.
- [6] AHMAD M J, TIWARI G N. Solar radiation models- a review[J]. International journal of energy research, 2011, 35(4): 271-290.
- [7] BUDYKO M I. The effect of solar radiation variations on the climate of the Earth [J]. Tellus, 1969, 21(5): 611-619.
- [8] 齐月,房世波,周文佐.近50年来中国东、西部地面太阳辐射变化及其与大气环境变化的关系[J].物理学报,2015,64(8):398-407.  
QI Y, FANG S B, ZHOU W Z. Correlative analysis between the changes of surface solar radiation and its relationship with air pollution, as well as meteorological factor in East and West China in recent 50 years[J]. Acta physica sinica, 2015, 64(8): 398-407.
- [9] 申彦波,赵宗慈,石广玉.地面太阳辐射的变化、影响因子及其可能的气候效应最新研究进展[J].地球科学进展,2008,23(9):915-923.  
SHEN Y B, ZHAO Z C, SHI G Y. The progress in variation of surface solar radiation, factors and probable climatic effects [J]. Advances in earth science, 2008, 23(9): 915-923.
- [10] 陶求华,李峥嵘,蒋福建.基于气象和空气质量观测数据的日太阳辐射估计[J].集美大学学报(自然科学版),2014,19(5):369-374.  
TAO Q H, LI Z R, JIANG F J. Estimation of incident daily solar radiation from meteorology and air quality observation[J]. Journal of Jimei University (natural science), 2014, 19(5): 369-374.
- [11] 石广玉,王标,张华,等.大气气溶胶的辐射与气候效应[J].大气科学,2008,32(4):826-840.  
SHI G Y, WANG B, ZHANG H, et al. The radiative and climatic effects of atmospheric aerosols[J]. Chinese journal of atmospheric sciences, 2008, 32(4): 826-840.
- [12] 王建源,冯建设,袁爱民.山东省太阳辐射的计算及其分布[J].气象科技,2006,34(1):98-101.  
WANG J Y, FENG J S, YUAN A M. Calculation and distributive characteristics of solar radiation in Shandong Province[J]. Meteorological science and technology, 2006, 34(1): 98-101.
- [13] MA J Y, LUO Y, SHEN Y B, et al. Solar energy forecasting methods and their applications and problems [J]. Resource science, 2011, 33(5): 829-837.
- [14] 马金玉.中国地面太阳辐射长期变化特征及短期预报方法研究[D].南京:南京信息工程大学,2011.  
MA J Y. The long-term variation and short-term prediction method of surface solar radiation in China [D]. Nanjing: Nanjing University of Information Science & Technology, 2011.
- [15] 高阳,张碧玲,毛京丽,等.基于机器学习的自适应光伏超短期出力预测模型[J].电网技术,2015,39(2):307-311.  
GAO Y, ZHANG B L, MAO J L, et al. Machine learning-based adaptive very-short-term forecast model for photovoltaic power[J]. Power system technology, 2015, 39(2): 307-311.

- [16] 王红睿. 机器学习背景下的太阳辐射数据基本重构方法 [C]//中国天文学会2018年学术年会摘要集. 昆明, 中国, 2018: 43.  
WANG H R. A fundamental reconstruction method for solar radiation data in the context of machine learning [C]// Abstracts of the 2018 Annual Meeting of the Chinese Astronomical Society. Kunming, China, 2018: 43.
- [17] 应王敏, 刘晓洁, 房世峰, 等. 基于机器学习的日尺度短波净辐射气候资源遥感反演研究 [J]. 资源科学, 2020, 42(10): 1998-2009.  
YING W M, LIU X J, FANG S F, et al. Retrieval of daily net surface shortwave radiation climatic resources based on machine learning [J]. Resources science, 2020, 42(10): 1998-2009.
- [18] VOYANT C, NOTTON G, KALOGIROU S, et al. Machine learning methods for solar radiation forecasting: a review [J]. Renewable energy, 2017, 105: 569-582.
- [19] AĞBULUT Ü, GÜREL A E, BIÇEN Y. Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison [J]. Renewable and sustainable energy reviews, 2021, 135: 110114.
- [20] ZHOU Y, LIU Y F, WANG D J, et al. A review on global solar radiation prediction with machine learning models in a comprehensive perspective [J]. Energy conversion and management, 2021, 235: 113960.
- [21] FALLAHI S, AMANOLLAHI J, TZANIS C G, et al. Estimating solar radiation using NOAA/AVHRR and ground measurement data [J]. Atmospheric research, 2018, 199: 93-102.
- [22] SUN H W, GUI D W, YAN B W, et al. Assessing the potential of random forest method for estimating solar radiation using air pollution index [J]. Energy conversion and management, 2016, 119: 121-129.
- [23] BENALI L, NOTTON G, FOUILLOY A, et al. Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components [J]. Renewable energy, 2019, 132: 871-884.
- [24] KARASU S, ALTAN A. Recognition model for solar radiation time series based on random forest with feature selection approach [C]//2019 11th International Conference on Electrical and Electronics Engineering (ELECO). Bursa, Turkey, 2020: 8-11.
- [25] ALSHARIF M, YOUNES M, KIM J. Time series ARIMA model for prediction of daily and monthly average global solar radiation: the case study of Seoul, South Korea [J]. Symmetry, 2019, 11(2): 240.
- [26] BOUALIT S, MELLIT A. SARIMA-SVM hybrid model for the prediction of daily global solar radiation time series [C]//2016 International Renewable and Sustainable Energy Conference (IRSEC). IEEE, Marrakech, Morocco, 2016: 712-717.
- [27] 丁一汇, 李巧萍, 柳艳菊, 等. 空气污染与气候变化 [J]. 气象, 2009, 35(3): 3-14, 129.  
DING Y H, LI Q P, LIU Y J, et al. Atmospheric aerosols, air pollution and climate change [J]. Meteorological monthly, 2009, 35(3): 3-14, 129.
- [28] 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(2): 159-166.
- [29] SAMANI Z. Estimating solar radiation and evapotranspiration using minimum climatological data [J]. Journal of irrigation and drainage engineering, 2000, 126(4): 265-267.
- [30] BAKIRCI K. Models of solar radiation with hours of bright sunshine: a review [J]. Renewable and sustainable energy reviews, 2009, 13(9): 2580-2588.
- [31] DÍAZ-TORRES J J, HERNÁNDEZ-MENA L, MURILLO-TOVAR M A, et al. Assessment of the modulation effect of rainfall on solar radiation availability at the Earth's surface [J]. Meteorological applications, 2017, 24(2): 180-190.
- [32] TANU M, AMPONSAH W, YAHAYA B, et al. Evaluation of global solar radiation, cloudiness index and sky view factor as potential indicators of Ghana's solar energy resource [J]. Scientific African, 2021, 14: e01061.
- [33] 张亮, 王赤, 傅绥燕. 太阳活动与全球气候变化 [J]. 空间科学学报, 2011, 31(5): 549-566.  
ZHANG L, WANG C, FU S Y. Solar variation and global climate change [J]. Chinese journal of space science, 2011, 31(5): 549-566.
- [34] KHODAKARAMI J, GHOBADI P. Urban pollution and solar radiation impacts [J]. Renewable and sustainable energy reviews, 2016, 57: 965-976.
- [35] 薛德强. 济南的城市发展对气候的影响 [J]. 气象, 1996, 22(2): 3-6.  
XUE D Q. Effects of development of Ji'nan City on climate [J]. Meteorological monthly, 1996, 22(2): 3-6.
- [36] 王建源, 赵玉金, 陈艳春, 等. 山东省太阳辐射及其光热生产潜力评估 [J]. 安徽农业科学, 2010, 38(7): 3581-3583, 3585.  
WANG J Y, ZHAO Y J, CHEN Y C, et al. Evaluation on solar radiation resource and photosynthetic and thermal potential productivity in Shandong Province [J]. Journal of Anhui agricultural sciences, 2010, 38(7): 3581-3583, 3585.
- [37] 桑博, 孙明虎. 济南市大气中SO<sub>2</sub>浓度变化特征研究 [J]. 中国环境管理干部学院学报, 2017, 27(4): 78-81.  
SANG B, SUN M H. A study on concentration variation of atmospheric SO<sub>2</sub> in Jinan [J]. Journal of Environmental

- Management College of China, 2017, 27(4): 78-81.
- [38] 曾贤刚, 阮芳芳, 姜艺婧. 中国臭氧污染的空间分布和健康效应[J]. 中国环境科学, 2019, 39(9): 4025-4032.
- ZENG X G, RUAN F F, JIANG Y J. Spatial distribution and health effects of ozone pollution in China[J]. China environmental science, 2019, 39(9): 4025-4032.
- [39] JAMALY M, KLEISSL J. Spatiotemporal interpolation and forecast of irradiance data using Kriging[J]. Solar energy, 2017, 158: 407-423.
- [40] 杨凌霄, 侯鲁健, 吕波, 等. 济南市大气细颗粒物水溶性组分及大气传输的研究[J]. 山东大学学报(工学版), 2007, 37(4): 98-103.
- YANG L X, HOU L J, LYU B, et al. Study on the water-soluble ions in fine particle matter and the long-range transport of air masses in the city of Ji'nan[J]. Journal of Shandong University (engineering science), 2007, 37(4): 98-103.
- [41] 沈志宝, 王尧奇, 季国良, 等. 太阳辐射在山谷城市污染大气中的削弱[J]. 高原气象, 1982, 1(4): 74-83.
- SHEN Z B, WANG Y Q, JI G L, et al. The attenuation of solar radiation in the polluted urban atmosphere[J]. Plateau meteorology, 1982, 1(4): 74-83.

## ANALYSIS OF MEDIUM- AND LONG-TERM CHANGES IN SOLAR RADIATION BASED ON RANDOM FOREST MODEL

Jia Xingbin<sup>1,2</sup>, Wang Guoju<sup>1,2</sup>, Wang Renzheng<sup>3</sup>, Gong Xiang<sup>1,2</sup>

(1. School of Mathematics and Science, Qingdao University of Science and Technology, Qingdao 266061, China;

2. Qingdao Innovation Centre for Artificial Intelligence, Qingdao 266061, China;

3. College of Environmental Science and Engineering, Ocean University of China, Qingdao 266100, China)

**Abstract:** This paper presents a comprehensive analysis of long-term solar radiation trends and influencing factors in Ji'nan City, Shandong Province, based on multi-source radiation observations, using the random forest (RF) algorithm, seasonal autoregressive integrated moving average (SARIMA) model and characteristic importance. The results show that the RF model fits the monthly solar radiation better, with the coefficient of determination and the mean absolute percentage error of 0.92 and 9%, respectively, which are better than the SARIMA model. The monthly solar radiation in Ji'nan City and the surrounding area experiences the process of "darkening" to "brightening" from 1980 to 2020. The maximum temperature and sunshine hours are the main factors affecting the accuracy of the monthly variation of solar radiation, rainfall is an important cause of sudden changes in total monthly solar radiation, and SO<sub>2</sub> and O<sub>3</sub> among atmospheric pollutants have the greatest correlation with solar radiation. This paper is important for guiding the development of the solar energy industry and environmental management in Ji'nan.

**Keywords:** solar radiation; random forest; feature selection; SARIMA model; fitting analysis