

基于TROPOMI的近地面臭氧浓度估算研究

张少磊, 杜宁*, 王莉, 吴宇宏, 吴迪, 周彬

贵州大学矿业学院, 贵州 贵阳

收稿日期: 2022年12月2日; 录用日期: 2022年12月16日; 发布日期: 2023年1月31日

摘要

在PM2.5浓度逐年下降的背景下, 臭氧浓度不降反升, 目前臭氧已经成为国内大气的主要污染物之一。本文利用当今大气监测中时空分辨率最高的TROPOMI传感器, 并基于大数据关联思想, 构建GWR-RF臭氧浓度估算模型, 融合臭氧浓度地面监测数据、欧洲中期天气预报中心的ERA5数据集、高分辨率遥感影像(TROPOMI_NO₂)数据, 构建训练整体数据集, 用以估算2019年四川省每日最大8 h平均臭氧浓度(O3_8h)。研究结果R²为0.94, RMSE为10.5 μg·m⁻³, MAE为7.49 μg·m⁻³, 表明GWR-RF模型对O3_8h有较好的估算性能。

关键词

近地面臭氧; 对流层观测仪(TROPOMI_NO₂); ERA5; GWR-RF

Estimation of Near-Surface Ozone Concentration Based on TROPOMI

Shaolei Zhang, Ning Du*, Li Wang, Yuhong Wu, Di Wu, Bin Zhou

School of Mining, Guizhou University, Guiyang Guizhou

Received: Dec. 2nd, 2022; accepted: Dec. 16th, 2022; published: Jan. 31st, 2023

Abstract

In the background of decreasing PM2.5 concentration year by year, ozone concentration does not fall on the contrary, ozone has become one of the main pollutants in the domestic atmosphere. In this paper, the TROPOMI sensor, which has the highest spatial and temporal resolution in current atmospheric monitoring, and based on the idea of big data association, is used to construct the

*通讯作者。

GCR-RF ozone concentration estimation model, which integrates the ground monitoring data of ozone concentration, the ERA5 dataset of the European Center for Medium Range Weather Prediction (ECMWF) and the high resolution remote sensing image (TROPOMI_NO₂) data. A training dataset was constructed to estimate the daily maximum 8h mean ozone concentration (O₃_8h) in Chengdu in 2019. The results showed that R², RMSE and MAE were 0.94, 10.5 μg·m⁻³ and 7.49 μg·m⁻³, indicating that the GWR-RF model had good performance in estimating O₃_8h.

Keywords

Near Surface Ozone, Troposphere Observatory (TROPOMI_NO₂), ERA5, GWR-RF

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前大气污染问题愈发严重, 已经由传统的化学成分燃烧引起的单一污染物现状逐渐转变为由多种污染物共同作用形成的复合型污染。臭氧(O₃)作为一种可以通过光学、化学反应等多种方式生成的二次污染物, 对人体和动植物的健康有着严重的危害。并且有研究表明人体长期暴露在高浓度的臭氧污染中则会降低心肺功能, 从而导致哮喘、心血管等多种疾病; 同时也有相关研究表明臭氧浓度过高则会抑制植物的正常生长。因此采用一种合理且有效的方法来准确评估臭氧浓度, 可以有效降低臭氧产生的危害。

进入 21 世纪以来, 越来越多的学者采用地面监测站点数据来研究多种大气污染成分浓度, 但是由于技术的限制、监测空间有限以及地面监测站点分布不均匀等原因, 这给大气污染研究带来了一些不确定性和不准确性。随着科学实力的提升, 多种航天遥感卫星技术高速发展, 遥感卫星所提供的监测数据为估算 PM_{2.5}、O₃、NO₂、SO₂ 等大气污染成分提供了可靠的支持[1] [2] [3]。但由于云覆盖等问题的影响, 导致卫星臭氧监测数据存在缺失的情况, 导致对地面臭氧浓度估算结果存在误差。

赵楠[4]采用 XGBoost 模型用来估算 2019 年每日近地面臭氧浓度, 研究结果表明 TROPOMI_NO₂ 与 O₃_8h 的相关性较强, 且可以对地面臭氧污染可以进行有效估算; 李一蜚[5]的研究利用 GBRT 算法, 对中国区域近地面臭氧浓度进行了估算, 其结果较为理想; Md Al MasumBhuiyan [6]的研究利用卫星遥感数据近地面臭氧浓度并对健康影响进行评价, 采用机器学习得到了不错的效果, 但是上述方法仍然有进步空间。

目前, 我国对臭氧(O₃)遥感探测的研究处于起步阶段, 存在研究案例少, 方法单一等问题, 且大多数集中在京津冀、长三角等[7] [8]地区, 针对于四川地区主要是采用地面监测站进行插值[9]并研究时空分布特征。本研究基于空间分辨率最高的 S-5P 卫星遥感数据, 并结合地面站点实测数据, 分析了空、地数据的相关性, 基于 GWR-RF 模型, 对四川省近地面臭氧(O₃)进行时空特征分析, 以期为今后量身定制城市空气污染防治策略提供参考。

2. 研究区与数据

2.1. 研究区域

本文选择的研究区域为四川省, 简称川或蜀, 省会成都。位于中国西南地区内陆, 地处长江上游,

素有“天府之国”的美誉。总面积约 48.6 万 km²，界于北纬 26°03′~34°19′，东经 97°21′~108°12′之间，与重庆、贵州、云南、西藏、青海、甘肃和陕西等 7 省(自治区、直辖市)接壤。

2.2. 研究数据

站点数据：地面监测站点数据来源于 2019 年 1 月 1 日到 2019 年 12 月 31 日的中国环境监测总站全国城市空气质量实时发布平台(<http://www.cnemc.cn/>)发布的逐小时数据。截止到 2019 年 1 月四川省总共布设有 105 个站点，剔除缺测、无效值等不符合标准后保留 93 个站点。本文按照《环境空气质量标准》(GB3095-2012)的规定[10]选取评价指标，日评价指标采用臭氧日最大 8h 滑动平均浓度(O3_8h)。月评价指标将每个月大于 27 个有效的 O3_8h 浓度值的站点和 2 月大于 25 个有效的 O3_8h 浓度值的站点计算每个站点的月均值。将臭氧日最大 8 h 滑动平均浓度的第 90 百分位数的浓度值(O3_8h_90)作为臭氧年评价指标。并依据《环境空气质量指数(AQI)技术规定(试行)》(HJ 633-2012) [11]，臭氧日最大 8 h 滑动平均浓度的一级限值为 100 μg·m⁻³，二级限值 160 μg·m⁻³，臭氧浓度超过 160 μg·m⁻³ 则认为超标。

在此说明点位 8 小时平均的计算方法，使用滑动平均的方式计算。对于指定时间 X 的 8 小时均值，定义为：X-7、X-6、X-5、X-4、X-3、X-2、X-1、X 时的 8 个 1 小时平均值的算术平均值，称为 X 时的 8 小时平均值。一个自然日内有 24 个点位 8 小时平均值，其时标分别记为 1:00、2:00、3:00、……、23:00 和 24:00 时。点位日最大 8 小时平均：点位一个自然日内 8:00 时至 24:00 时的所有 8 小时滑动平均浓度中的最大值。

TROPOMI 数据：欧洲航天局(ESA)于 2017 年 10 月 13 日发射了 Sentinel-5P 卫星，该卫星主要用于对空气质量、气候作用力、臭氧和紫外辐射等方面进行高分辨率测量。TROPOMI 是 ESA/Copernicus Sentinel-5 Precursor 任务的有效载荷，在低地球太阳同步极地轨道上运行，轨道宽达 2600km，每天覆盖全球。

目前 TROPOMI 数据可以在(<https://s5phub.copernicus.eu/dhus/>)免费下载，TROPOMI 数据下载以后采用 netcdf4 格式存储，并提供 L1B 和 L2 两种数据产品，产品分别为 Offline、Near Real Time、Reprocessing 三种级别。本文采用的是 2019 年 TROPOMI 的 L2 级数据产品(NO₂ 浓度数据，单位为 mol/m²)，在 S5P 产品中提供了质量保证值(qa)，以使用户剔除不准确的观测值，官方推荐使用 qa 值大于 0.75 的数据，以消除多云、冰雪覆盖等其他问题的干扰。

气象数据：气象因子可以提高模型估算的准确性和可靠性。本文采用 ECMWF 提供的 ERA5 数据集，相比于 EAR-Interim 的数据集，ERA5 气象再分析资料增加了新的参数，提高了数据的准确性，且具有较高时空分辨率和时间跨度。本文选取的变量包括边界层高度(BLH)、低植被覆盖率(CVL)、地表气压(SP)、2 m 露点温度(D2M)、2 m 地面温度(T2M)、10 m 经向风(V10)、10 m 纬向风(U10)、高植被覆盖指数 CVH、总降雨(RAIN)，晴空地表净辐射(SSRC)。

3. 方法

地理加权回归(Geographically Weighted Regression, GWR)是一种空间分析技术，广泛应用于地理学及涉及空间模式分析的相关学科。GWR 通过建立空间范围内每个点处的局部回归方程，来探索研究对象在某一尺度下的空间变化及相关驱动因素，并可用于对未来结果的预测。由于它考虑到了空间对象的局部效应，因此其优势是具有更高的准确性。

随机森林(Random Forests)指的是利用多棵树对样本进行训练并预测的一种分类器。RF 由 Breiman 提出，可用于分类和回归。RF 由决策树组成，其通过从原训练集中有放回的随机抽取原训练集的子集输入决策树建模，可从每颗决策树获取一次预测结果，将所有决策树的预测结果求平均值即为随机森林

的预测结果。与其他机器学习算法相比，随机森林算法能处理高维度数据，并且无需做特征选择，具有对数据集的适应力强、实现简单、精度高和抗过拟合能力强等优势。RF 算法示意图见图 1。

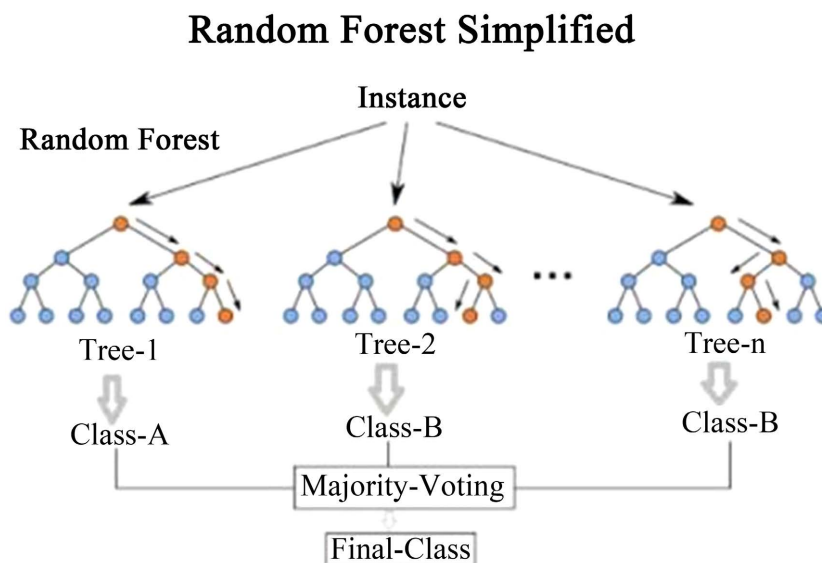


Figure 1. RF diagram
图 1. RF 示意图

模型评估指标，为客观评价模型的估算精度，采用决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)来比较不同模型拟合效果。 R^2 是度量多元回归中拟合程度的一个统计量，反映了在因变量 y 的变差中被估计的回归方程所解释的比例。 R^2 越接近 1，表明回归平方和占总平方和的比例越大，拟合程度就越好；均方根误差(RMSE)和平均绝对误差(MAE)则越小越好。计算见式(3.1、3.2、3.3)：

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_3^r(i) - O_3^p(i))^2}{\sum_{i=1}^N (O_3^r(i) - \bar{O}_3^r)^2} \quad (3.1)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_3^r(i) - O_3^p(i))^2} \quad (3.2)$$

$$\text{MAE} = \frac{1}{N} \left| (O_3^r(i) - O_3^p(i)) \right| \quad (3.3)$$

式中， N 为样本总数； O_3^r 为实测臭氧浓度； O_3^p 为估算臭氧浓度； \bar{O}_3^r 为实测臭氧浓度的平均值。

4. 结果与分析

4.1. 模型估算性能分析

研究区 2019 年 1 月~2019 年 12 月模型估算 O_3 _8h 的拟合结果如图 2 所示，结果表明模型 GWR-RF 模型估算结果与地面站点监测结果吻合度较高， R^2 为 0.94，RMSE 为 $10.5 \mu\text{g}\cdot\text{m}^{-3}$ ，MAE 为 $7.49 \mu\text{g}\cdot\text{m}^{-3}$ 。表明 GWR-RF 模型能很好的捕捉 O_3 _8h 与各影响因素之间复杂的非线性关系，模型拟合精度显著。估算结果较为可靠。可以充分证明该模型具有更好的模拟效果，估算的 O_3 _8h 浓度更接近真实值，可用于获取更高精度的四川省 O_3 _8h 时空分布特征，为区域空气质量评估和人体暴露风险评价及环境污染治理提供更加合理的科学参考。拟合结果见图 2 所示。

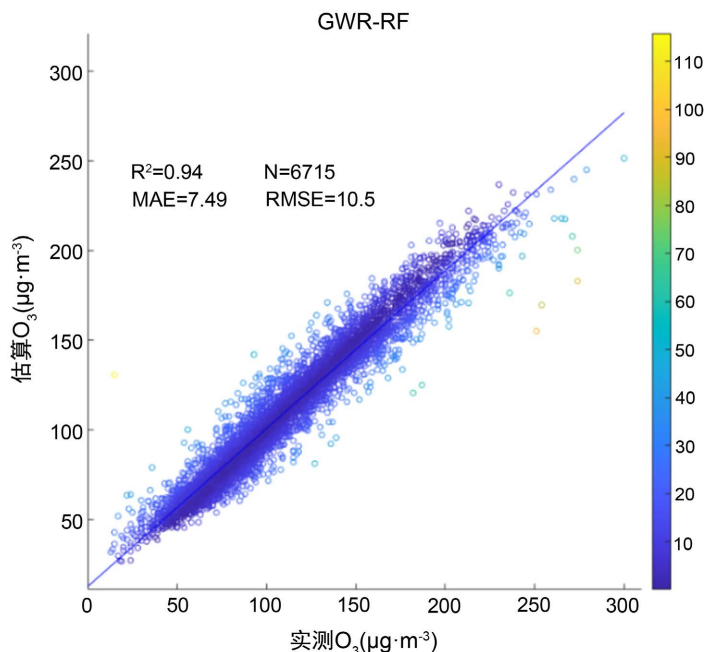


Figure 2. Fitting results
图 2. 拟合结果

4.2. 估算结果分析

图 3 是 GWR-RF 模型估算的 2019 年臭氧浓度分布图。从估算结果可以看出在四川省境内，臭氧的分布特征为东高西低，且东部地区的污染程度远远高于西部地区，导致这种分布的原因是因为包括成都市、泸州市、宜宾市等在内的多个城市人口数较多，且各种工业设施以及汽车尾气排放量较大。全省地形复杂地貌多样，地貌类型为山地、丘陵、平原和高原 4 种类型，四川地形特征体现为西高东低，由西北向西南倾斜的特点，东部和北部丘陵为主，西部以高原为主。因四川省独特的地理位置和气候条件，风向受反气旋控制明显，易形成明显的“回水区”或静风现象，不利于污染物扩散，污染气体容易在盆地底部聚集，形成典型的地形逆温现象，造成污染物不断累积，且不易流通，使得四川省中东部成为大气污染较为严重的地区。除此之外重庆市虽然未在本次研究区域内，但因为臭氧气体极易受环境(如风速、风向、温度，气压等)影响，因此重庆市产生的大量污染气体也对盆地地区的臭氧污染存在一定的贡献程度。

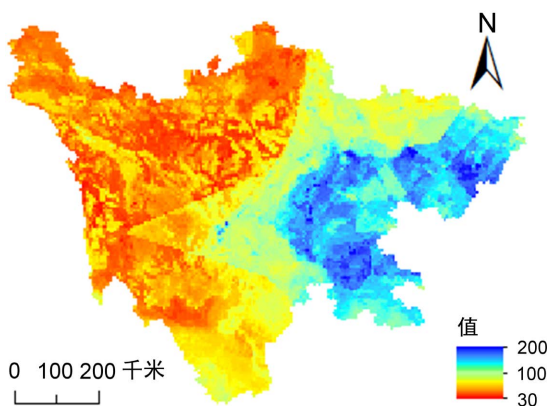


Figure 3. Ozone estimates
图 3. 臭氧估算图

5. 结论

本文在空气质量监测网络的支持下, 结合大数据关联的思想, 以四川省为研究目标, 通过构建 GWR-RF 模型来估算 2019 年四川省的每日近地面臭氧浓度。在估算过程中不仅对研究的影响因子进行地理加权回归计算, 并将估算结果在进行机器学习拟合, 使得 GWR-RF 模型在预测方面表现出较高的性能。在 GWR-RF 算法较高性能的表现下, 使用高时空分辨率遥感影像 TROPOMI_NO₂ 可以提高模型估算的准确性。在今后可以加入各种影响因子并采用该模型进行研究区域的臭氧浓度估算, 最后可以得出研究区域臭氧浓度污染的分布情况, 并分析主要污染源, 为空气治理提供相关建议等。

参考文献

- [1] 李旭文, 张悦, 姜晟, 王甜甜, 纪轩禹, 茅晶晶, 丁铭. “哨兵-5P”卫星 TROPOMI 传感器在江苏省域大气污染监测中的初步应用[J]. 环境监控与预警, 2019, 11(2): 10-16.
- [2] 高诚, 张超, 余树全. 2005-2013 年长江三角洲地区对流层二氧化氮时空变化特征[J]. 浙江农林大学学报, 2015, 32(5): 691-700.
- [3] 周亮, 周成虎, 杨帆, 王波, 孙东琪. 2000-2011 年中国 PM_{2.5} 时空演化特征及驱动因素解析[J]. 地理学报, 2017, 72(11): 2079-2092.
- [4] 赵楠, 卢毅敏. 基于 XGBoost 算法的近地面臭氧浓度遥感估算[J]. 环境科学学报, 2022, 42(5): 95-108.
- [5] 李一蜚, 秦凯, 李丁, 等. 基于梯度提升回归树算法的地面臭氧浓度估算[J]. 中国环境科学, 2020, 40(3): 997-1007.
- [6] Al Masum, B.M., *et al.* (2021) Machine Learning Techniques Applied to Predict Tropospheric Ozone in a Semi-Arid Climate Region. *Mathematics*, **9**, 2901.
- [7] 姚青, 马志强, 郝天依, 等. 京津冀区域臭氧时空分布特征及其背景浓度估算[J]. 中国环境科学, 2021, 41(11): 4999-5008.
- [8] 黄小刚, 邵天杰, 赵景波, 等. 长三角城市群臭氧浓度的时空分异及驱动因素[J]. 长江流域资源与环境, 2019, 28(6): 1434-1445.
- [9] 卢宁生, 张小玲, 杜云松, 等. 成都平原城市群夏季臭氧污染天气形势与潜在源分析[J]. 高原气象, 2022: 1-14.
- [10] 中华人民共和国国家标准: 环境保护部. GB 3095-2012 环境空气质量标准[S]. 北京: 中国环境科学研究院, 2018.
- [11] 中华人民共和国国家环境保护标准: 环境保护部. HJ 633-2012 环境空气质量指数(AQI)技术规定(试行) [S]. 北京: 中国环境科学出版社, 2012.