#### doi: 10.6046/zrzyyg.2023220

第37卷,第1期

2025年02月

引用格式:陈薪,施国萍. 基于机器学习的 FY-4A 气溶胶光学厚度反演[J].自然资源遥感,2025,37(1):213-220.(Chen X, Shi G P.Machine learning-based inversion of aerosol optical depth inversion from FY-4A data[J]. Remote Sensing for Natural Resources,2025,37(1):213-220.)

# 基于机器学习的 FY-4A 气溶胶光学厚度反演

# 陈 薪, 施国萍

(南京信息工程大学地理科学学院,南京 210044)

**摘要**: 我国新一代静止气象卫星 FY-4A 搭载了先进静止轨道辐射成像仪(advanced geostationary radiation imager, AGRI),为了针对 AGRI 数据的特点开发基于机器学习算法的高时空分辨率 AGRI 气溶胶数据集,利用 2021 年中国 及周边地区 67 个全球地基气溶胶自动观测网络系统(aerosol robotic network, AERONET)站点数据,选取 AGRI 数据 中表观反射率数据、观测角度数据、高程和 MODIS 的地表反射率等因子,采用随机森林(random forest, RF)、梯度增 强回归(gradient boosting regression tree, GBRT)、极端梯度增强(extrme gradient boosting, XGBoost)和反向传播神经网 络(back propagation neural network, BPNN)4 种机器学习方法反演气溶胶光学厚度(aerosol optical depth, AOD)。根 据模型评价指标选择最优模型,实现基于 FY-4A 的 AOD 反演,空间分辨率 4 km,并与时刻相近的 MODIS 气溶胶 产品进行对比分析。结果表明: 4 种机器学习建立的 AOD 反演模型相关系数 R 均在 0.90 以上,平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE)均在 0.09 和 0.14 以下,机器学习模型对 AOD 反演具有一定的可行性; GBRT 模型在 4 种机器学习中反演精度最优,验证误差中 R 为 0.82, MAE 为 0.16, RMSE 为 0.25, 47%的反演结果落在期望误差内,表明 GBRT 反演出的 FY-4A AOD 反演信相对 MODIS AOD 在空间分布上具有较好的一致性,83.57%的网格偏差集中在[-1.0,0)之间,FY-4A AOD 反演值相对 MODIS AOD 略高。**关键词**: 机器学习; 气溶胶光学厚度; FY-4A/AGRI 数据; AERONET; MODIS AOD 产品

中图法分类号: TP 79; X 513 文献标志码: A 文章编号: 2097-034X(2025)01-0213-08

# 0 引言

气溶胶光学厚度(aerosol optical depth, AOD)作 为气溶胶最重要的参数之一, 是表征大气浑浊的关 键物理量, 也是确定气溶胶气候效应的重要因素。 通常, 高 AOD 值预示着气溶胶纵向积累的增长, 导 致了大气能见度的降低。大气气溶胶在气候变 化<sup>[1]</sup>、大气环境监测<sup>[2]</sup>以及人类健康<sup>[3]</sup>等领域均有 重要的应用。因此, 对于 AOD 的估算反演是气溶胶 监测的重要研究工作之一。

地面观测可以获得准确和连续的气溶胶信息, 但由于站点分布稀疏,无法从细节上展现气溶胶的 空间分布信息,而卫星遥感在大气环境质量变化的 连续性、空间性和趋势性监测方面具有明显优势。 气溶胶反演算法有 MODIS 的暗目标(dark target algorithm, DT)算法<sup>[4-5]</sup>和深蓝(deep blue algorithm,

DB)算法<sup>[6-7]</sup>等。2022年, Chen 等<sup>[8]</sup>统计了 2012— 2019 年 MODIS DT 与地基 AOD 的相关系数 R 和均 方根误差(root mean square error, RMSE), 分别为 0.820和 0.157, MODIS DB 的 R 和 RMSE 分别为 0.827和 0.134, MODIS DT 与 DB 的融合产品 DTB 的 R和RMSE分别为0.853和0.125。日本Himawari-8/ AHI提供了东亚上空 AOD 产品(V2.1 版本)与 AERONET 站点观测值的 R 达到 0.67, 平均绝对误 差(mean absolute error, MAE)为 0.19, RMSE 为 0.37<sup>[9]</sup>。我国风云系列卫星中,多个搭载探测气溶 胶传感器的卫星成功发射,推动了基于风云气象卫 星数据反演气溶胶的研究发展。研究人员先后尝试 利用 FY-1C/1D 数据(极轨卫星)<sup>[10]</sup>、FY-2C/VIS-SR 数据(静止卫星)<sup>[11]</sup>、FY-3A/B MERSI 数据(极 轨卫星)<sup>[12]</sup>进行 AOD 反演算法构建,其中刘智超 等<sup>[12]</sup>研究发现 FY-3A 的反演结果与太阳光度计观 测值的 R 高于 0.96, 仅有 20%的样本表现出较大的

收稿日期: 2023-07-20;修订日期: 2024-01-04

**基金项目:** 国家自然科学基金青年项目"基于 SUNFLUX 辐射参数化计算方案的起伏地形云天实际地表太阳辐射分布式模拟研究及其在陆面过程中的应用"(编号: 41805083)资助。

**第一作者:**陈 薪(1999-),女,硕士研究生,主要从事 3S 集成与气象应用研究。Email: cx15151838939@163.com。

通信作者:施国萍(1984-),女,博士,副教授,主要从事 3S 集成与气象应用研究。Email: shiguopingnj@163.com。

偏差,FY-3B的 R 最高为 0.77, RMSE 最大为 0.35。 2016年,中国第二代地球静止气象卫星 FY-4A 成 功发射。FY-4A 上搭载的先进静止轨道辐射成像 仪(advanced geostationary radiation imager, AGRI)可 实现分钟级的区域快速扫描,高频次获取14波段的 地球云图,因此FY-4A/AGRI在大气污染监测和建 模方面具有很大的潜力<sup>[13]</sup>。许梦婕<sup>[14]</sup>基于 FY-4A/AGRI 数据,采用 DT 和 DB 2 种传统算法反演京 津冀地区 AOD, R<sup>2</sup>分别为 0.755 和 0.607; Jiang 等<sup>[15]</sup>针对 FY-4A 数据提出一种新的陆地 AOD 反 演算法----BRL 算法, FY-4A AOD, MODIS DT AOD 和 MODIS DB AOD 与地基观测值的 R 均为0.84, FY-4A AOD 的 RMSE(0.16) 低于 MODIS DT AOD (0.20),高于 MODIS DB AOD(0.12),说明反演结果 优于 MODIS DT AOD 产品,略差于 MODIS DB AOD 产品。近年来,随着机器学习的广泛应用,为气溶胶 遥感带来了新的研究和应用手段<sup>[16]</sup>。Ristovski 等[17]利用 AERONET 地基气溶胶观测的 AOD 和 MODIS 卫星遥感数据样本训练神经网络模型, AOD 测试精度(R<sup>2</sup>=0.819, RMSE=0.100)优于当时的 DT 方法 CO05 版本产品(R<sup>2</sup>=0.786, RMSE=0.108)<sup>[18]</sup>; Ding 等<sup>[19]</sup>利用全卷积神经网络反演的 FY-4A AOD 与地面观测值的 MAE 和 RMSE 分别达到 0.145 和 0.237。机器学习是通过学习样本数据获得复杂方 程精确解的有力工具<sup>[20]</sup>,具有强大的非线性拟合能 力,对于定量机理难以精确描述的定量化参数反演 问题,可以作为一种统计方法实现遥感信息提取。

本文选用 2021 年 FY-4A 上搭载的 AGRI 中的 4 km 光谱通道和产品集,经过波长插值与时间匹 配,构建样本数据集。基于机器学习算法构建随机 森林(random forest, RF)、梯度增强回归(gradient boosting regression tree, GBRT)、极端梯度增强(extrme gradient boosting, XGBoost)和反向传播神经网 络(back propagation neural network, BPNN)4种算法 反演模型。模型的反演结果与全球地基气溶胶自动 观测网络系统(aerosol robotic network, AERONET) 站点观测的 AOD 进行检验,对比各方法之间的准确 性和稳定性,选择出较优的机器学习模型。将较优 的机器学习模型反演所得的 AOD 结果与相近时刻 的 MODIS 气溶胶数据进行对比,探讨机器学习方法 用于气溶胶动态监测的可行性与准确性。

1 研究区概况及数据源

# 1.1 研究区概况

研究区范围为 E65°~145°,N0°~55°(图1)。该 研究区具有复杂的地理环境和差异化的气象气候条 件。由于东亚及东南亚地区经济发展迅速和人口密 集,空气污染严重,气溶胶对空气质量的影响备受关注。



图 1 研究区及站点分布 Fig.1 Study area and site distribution

1.2 数据源及其预处理

1.2.1 FY-4A/AGRI 遥感数据

本文选取 AGRI 中 12 个表观反射率数据(表 1),成像时刻的太阳天顶角(solar zenith angle,

Tab.1 AGRI channels								
通道序号	通道类型	中心波长/μm	光谱带宽/μm	主要用途				
B1		0.47	0.45~0.49	气溶胶				
B2	可见光与近红外	0.65	0.55~0.75	植被				
B3		0.825	0.75~0.90	植被				
B4	1	1.375	1.36~1.39	卷云				
В5	短波红外	1.61	1.58~1.64	云、雪				
B6		2.25	2.1~2.35	卷云、气溶胶				
B7	中連行从	3.75	3.5~4.0(高)	火点				
B8	中被红外	3.75	3.5~4.0(低)	地表				
В9	十次	6.25	5.8~6.7	云导风				
B10	小八	7.1	6.9~7.3	云导风				
B11	长速红旗	8.5	8.0~9.0	云导风、云				
B12	下放红外	13.5	13.2~13.8	云顶高度				

表1 AGRI 数据波段

SOZ)、太阳方位角(solar azimuth angle, SOA)、太阳 耀斑角(solar flare angle, SGA)和卫星天顶角(satellite zenith angle, SAZ)以及选取晴朗像元的云检测 产品值用于建立训练模型。

1.2.2 MODIS 产品

本文选取 MOD09A1 第 3 波段( 蓝光) 地表反射 率参与模型的训练, 空间分辨率为 500 m, 时间分辨 率为 8 d。使用 MODIS DT 和 DB 算法融合 0.55 μm 处的 AOD 产品( MODIS DTB), 空间分辨率为 10 km, 时间分辨率为 0.5~1 d, 该数据集精度优于前一代 的 C5 产品且优于 DT 或 DB 产品<sup>[8]</sup>。

1.2.3 AERONET 地基观测数据及处理

选取研究区内的 67 个 AERONET 站点实测数 据(Level1.5 和 Level2.0),为训练样本集提供精确 AOD 信息,并应用于 AOD 反演的训练与验证。站 点分布如图 1 所示。由于 AGRI 的扫描周期为 15 min,因此本文对 AERONET 每个站点在相应 15 min 内的数据作平均,与 AGRI 数据对应,建立训练集。

AERONET AOD 产品包含 1.02, 0.87, 0.675, 0.5, 0.44, 0.38 和 0.34 μm 波段的 AOD, 而 MODIS 只 包含 0.47, 0.55 和 0.66 μm 波段的 AOD 产品, 以及 AGRI 反演结果为 0.55 μm 波段的 AOD。因此, 需 要对 AERONET 数据进行波长插值, 得到与卫星产 品波段一致的 AOD。相对于 Angstrom 波长指数插 值方法, 二次多项式插值法可插值出精度较高、更可 靠的 0.55 μm 通道处 AOD 值<sup>[21]</sup>。本研究采用二次 多项式插值法进行波段插值, 选取 AERONET 的 0.675, 0.5 和 0.44 μm 3 个波段, 进行插值获取 0.55 μm 的 AOD。插值公式为:

$$\ln \tau_{\lambda} = a_0 + a_1 \ln \lambda + a_2 \left( \ln \lambda \right)^2 , \qquad (1)$$

式中:  $\tau_{\lambda}$  为  $\lambda$  通道处的 AOD 值;  $a_i$  (i = 0, 1, 2)为 未知系数。

1.2.4 高程数据

高程数据为通过地理空间数据云(https://www.gscloud.cn/)获取到的90m空间分辨率的SRTM3数据。

2 研究方法

#### 2.1 机器学习建模

本文使用基于 Bagging 方法的 RF 模型、基于 Boosting 方法的 XGBoost 和 GBRT 模型,以及基于深 度学习的 BPNN 模型 4 种机器学习方法反演 AOD。

将 67 个站点 2021 年 1—12 月的数据作为训练 集,2022 年 7 月、10 月和 2023 年 1 月、4 月作为验证 集。训练集参与反演模型建立、训练以及采用五折 交叉验证方法进行参数调优,共匹配到 69 723 组样 本。验证集不参与训练,作为独立样本(18 300 组) 对经过参数调优后的模型进行稳定性检验。

RF,XGBoost 和 GBRT 的调参过程是通过网格 搜索(GridSearchCV)对不同的参数选择和不同参数 取值的组合进行尝试,调整的参数主要包括决策树 个数(n\_estimators)和决策树最大深度(max\_depth) 等超参数。如 n\_estimators 从 0~2 000, max\_depth 从 0~100 分别间隔 20 进行迭代,也就是 1 000 种参 数组合。BPNN 模型的调参过程是不断调整隐含层 层数、神经元数、迭代次数等,隐含层的最佳神经元 数可在  $2\sqrt{m} + \mu 和 2m + 1 之间确定,其中 m 和 \mu 分$ 别为输入层和输出层的神经元数<sup>[22]</sup>,本文取 m = $18, <math>\mu = 1$ ,也就是在 9~37 范围内确定隐含层的神经 元数。4 种机器学习模型通过训练集采用五折交叉 验证方法寻找到最优的模型参数组合。

#### 2.2 重要性方法

计算特征重要性主要是计算每一个特征贡献的 平均值,然后将其特征做对比<sup>[23]</sup>。通常采用袋外数 据(out of bag,OOB)的错误率来辅助筛选,计算公 式为:

$$FIM_{j} = \frac{\sum (err \ OOB_{2} - err OOB_{1})}{N} \quad , \quad (2)$$

式中: *FIM* 为特征的重要性评分; *j* 为特征序数; *N* 为决策树个数; *errOOB*<sub>1</sub> 为正常袋外误差; *err OOB*<sub>2</sub> 为有噪声扰乱时的袋外误差。

# 2.3 验证指标

本研究使用4个验证指标来表征反演结果与真 实值之间的差异,分别是:R,MAE,RMSE和期望误 差(expected error,EE)。其中为更好地评定 FY-4A/AGRI 气溶胶反演的精度,选择得到散点落入 EE 范围之内、之上、之下的百分比(=EE,>EE, <EE)作为精度衡量参数<sup>[24-26]</sup>,分析反演结果与真 实值的整体分布情况,各指标公式为:

$$R = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} , \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_i| \quad , \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i)^2} , \qquad (5)$$

 $EE = \pm (0.05 + 0.15 Y_i) , \qquad (6)$ 

式中: $X_i$ 为遥感 AOD 反演值; $Y_i$ 为地基 AOD 实测

值; X和Y分别为反演值和实测值的平均值; n为地 基实测数据与遥感数据有效匹配结果的数量。

# 3 结果与分析

# 3.1 特征重要性

特征选择是构建机器学习模型的重要步骤之 一,旨在通过使用相关特征来减少机器学习模型的 输入特征,并去除数据集中的噪声。在模型构建中 选择相关特征的因子会直接影响模型的准确性和复 杂性<sup>[27]</sup>。本文将 FY-4A/AGRI 12 个波段(B1-B12)的表观反射率,地表反射率(MOD09A1),SOZ, SOA, SGA, SAZ 和高程(Elevation) 作为 18 个输入变 量,分析特征重要性(图2)。由图2可知,模型变量 的排序从大到小前5位为:SAZ,B1,Elevation,B3 和 SOZ,其中前 3 个变量对 AOD 估计值的贡献约 52%。SAZ和 SOZ 会影响气溶胶粒子对太阳光的吸 收和散射作用,所以,观测角度数据是反演 AOD 的 重要变量; B1(蓝光波段)的中心波长为 0.47 μm, 与真值波段 0.55 μm 相近, 而且蓝光波段对气溶胶 相当敏感,很容易受到其影响; B3(近红外波段)的 中心波长为 0.825 µm,用于监测植被作物长势和植 物分类等,因此 AOD 可能受植被因素的影响;高程 对 AOD 的影响主要是通过大气压力和温度的变化 来实现的,随着高程的增加,大气压力和温度均会下降,从而对 AOD 产生影响<sup>[28]</sup>。



Fig.2 Importance of features

#### 3.2 建模误差

以 AERONET 地基观测数据为真值,利用 4 种 机器学习方法建立基于 FY-4A 的 AOD 反演模型。 建模误差结果如图 3 所示,蓝线为 EE;黑线为 y= x;红线为反演值与真值的拟合线。4 种机器学习模 型的 R 均高于 0.90, *MAE* 小于 0.09, *RMSE* 小于 0.14。其中,GBRT 模型精度最优,*R*, *MAE* 和 *RMSE* 分别为 0.936 7,0.072 9 和 0.107 9,其次为 XGBoost 与 RF 模型,两者精度相当, BPNN 模型次之。



Fig.3 Scatterplot of modeling results for AOD and ground-based data for BPNN, RF, XGBoost and GBRT model

# 3.3 验证误差

第1期

#### 3.3.1 总体验证

选用与建模不同的时间数据(2022 年 7 月、10 月和 2023 年 1 月、4 月)代入已建立的 4 种机器学 习模型中,通过验证指标分析各模型的时间稳定性。 如图 4 所示,4 种机器学习模型的 *R*,*MAE*,*RMSE* 分 别在 0.76~0.82,0.15~0.17,0.24~0.26 范围内,约 44%及以上的反演结果落在 EE 内。整体上, 拟合线的斜率均小于 1, 较大的 AOD 会出现低估现象。 其中, GBRT 模型的 *R*, *MAE* 和 *RMSE* 分别约为0.82, 0.16, 0.25, 47%及以上的反演结果落在 EE 内, 说明 GBRT 模型较其他 3 种模型在时间稳定性上具有一定的优势。



图 4 BPNN, RF, XGBoost 和 GBRT 模型反演的 AOD 与地基数据的验证结果散点图

Fig.4 Scatterplot of validation results for AOD and ground-based data for BPNN, RF, XGBoost and GBRT model

3.3.2 分月验证

表2显示了使用 GBRT 模型对4个月份 FY-4A/AGRI 反演 AOD 的验证结果。4个验证月份的 R均在0.74以上,其中2023年1月R能够达到 0.90,说明反演结果在不同月份上具有较高的精度。 2022年7月、10月和2023年1月的 MAE 均在0.13 以下, RMSE 均在0.2以下,2023年4月的验证精度 有所下降。通过比较反演值与真实值,发现偏差较 大的情况较多出现在泰国站点(Chiang\_Mai\_Met\_ Sta)和印度站点(Dibrugarh\_Univ)。泰国的新年在 每年的4月份举办,人类活动较为密集,产生更多的 污染气体和粉尘<sup>[29]</sup>;另外,2023年4月印度发生极 端高温事件,温度与AOD显著正相关,进而促进次

	表 2	分月	月验证表	
Tab.2	Mon	thly	validation	table

时间	R	MAE	RMSE	= <i>EE</i> /%	> <i>EE</i> /%	< <i>EE</i> /%
2022年7月	0.748 8	0.120 6	0.188 6	52	20	28
2022 年 10 月	0.826 9	0.117 8	0.182 8	56	12	32
2023年1月	0.903 7	0.125 3	0.197 8	51	6	43
2023年4月	0.802 2	0.251 3	0.364 7	32	2	66

生有机气溶胶转化<sup>[30]</sup>。因此,可能造成反演偏差稍 大的原因是 *AOD*>1.5 的样本较少,对高 *AOD* 值的 训练不足,导致 FY-4A AOD 值出现低估现象。

# 3.4 空间分布对比

通过建模误差和时间稳定性验证,GBRT方法 在反演 AOD 中的性能要优于 XGBoost,RF,BPNN。 因此,使用 GBRT 模型来反演研究区的 AOD 空间分 布图。图 5(a)—(h)为 2022 年 7 月 4 日、2022 年 10 月 8 日、2023 年 1 月 25 日、2023 年 4 月 15 日 MODIS/Terra(UTC02:30)和 FY-4A/AGRI(UTC02: 45)空间分布特征对比图。图 5(i)—(p)为相同日 期 MODIS/Aqua (UTC05:30)和 FY - 4A/AGRI (UTC05:45)空间分布特征对比图。由于云的遮挡, FY-4A 反演结果与 MODIS 产品图像上都存在一定 的缺失。从整体分布来看,FY-4A AOD 与 Terra/ MODIS 和 Aqua/MODIS AOD 在空间趋势上较为一 致。图上的 4 d高值区域主要集中在中国的黄淮海 平原和印度河—恒河平原地区。黄淮海平原地区 AOD 较高主要是因为经济的快速发展和人口的密 集分布导致污染物的大量排放<sup>[31]</sup>;印度河—恒河 平原的部分地区出现高 AOD 的原因是当地的人为

排放和来自印度西北部塔尔沙漠的远程自然粉尘气 溶胶排放<sup>[32]</sup>。



图 5 MODIS/Terra, MODIS/Aqua 和 FY-4A/AGRI 产品区域分布对比 Fig.5 Comparison of the regional distribution of MODIS/Terra, Aqua/MODIS and FY-4A/AGRI products

为了进一步定量分析两者产品差值分布差异, 将 MODIS AOD 重采样至 4 km 的空间分辨率,再与 FY-4A AOD 相减。计算每组差值的频数与样本总 数的比例,得到差值频率直方图(图 6)。由图可见,



图 6 MODIS AOD 与 FY-4A AOD 的频率差异直方图 Fig.6 Histogram of frequency differences between MODIS AOD and FY-4A AOD

两者个别网格差值达到±2.5。总体误差范围在 [-1.0,0.2),占比达到93.07%,表明 MODIS AOD 与 FY-4A AOD 产品存在一定的一致性。当范围为 [-1.0,0)时,占比达到83.57%,说明FY-4A AOD 反演值相对 MODIS AOD 略高。最大频率主要集中 在[-0.2,-0.1)范围内,占比达到24.32%。当 AOD 差值>1 或<-1时,占比和为1.4%,FY-4A AOD 与 MODIS AOD 在较大的偏差上分布较少。

# 4 结论

利用 RF, GBRT、XGBoost 和 BPNN 这 4 种机器 学习方法,基于 2021 年 FY-4A 卫星对 AOD 进行反 演,分析模型输入参数对 AOD 反演结果的影响,并 将反演结果与 AERONET 地基观测数据结果和 MO-DIS 气溶胶产品进行比较验证,得出以下结论: 1) 基于 FY-4A/AGRI 数据,4 种机器学习方法 建模训练中反演值与站点观测值的 R, MAE, RMSE 分别在 0.90 以上,0.09 以下,0.14 以下,反演值与站 点观测值的相关性较高且误差较小。说明利用机器 学习反演 AOD 具有一定的可行性。

2) 独立验证时,4 种方法与站点观测值的 R 均 在 0.76 以上, MAE 和 RMSE 分别在 0.15 和 0.26 以 下,至少 44%的反演结果落在 EE 内。GBRT 时间稳 定性最高, R 达到 0.82, MAE 和 RMSE 分别降低到 0.16和 0.25,约 47%的反演结果落在 EE 内。BPNN, XGBoost 和 RF 精度次之。

3) FY-4A AOD 与 MODIS AOD 产品相比,两者 的空间分布特征与趋势相一致。两者之间的 AOD 偏差主要分布在[-1.0,0)之间,占比达到 83.57%, 表明 MODIS AOD 与 FY-4A AOD 反演结果具有较 好的一致性,且 FY-4A AOD 大多高于 MODIS AOD。

# 参考文献(References):

- Mishra M K, Rathore P S, Misra A, et al. Atmospheric correction of multispectral VNIR remote sensing data: Algorithm and inter-sensor comparison of Aerosol and surface reflectance products [J]. Earth and Space Science, 2020, 7(9):e2019EA000710.
- [2] Xie Y, Xue Y, Guang J, et al. Deriving a global and hourly data set of aerosol optical depth over land using data from four geostationary satellites: GOES-16, MSG-1, MSG-4, and himawari-8[J].IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(3): 1538-1549.
- [3] Sosnowski T R. Aerosols and human health: A multiscale problem[J].Chemical Engineering Science, 2023, 268:118407.
- [4] Kaufman Y J, Wald A E, Remer L A, et al. The MODIS 2.1-/ spl mu/m channel-correlation with visible reflectance for use in remote sensing of aerosol[J].IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(5):1286-1298.
- [5] Levy R C, Remer L A, Mattoo S, et al.Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance [J]. Journal of Geophysical Research: Atmospheres, 2007, 112(D13):211-231.
- [6] Hsu N C, Tsay S C, King M D, et al. Aerosol properties over brightreflecting source regions [J].IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(3):557-569.
- [7] Hsu N C, Tsay S C, King M D, et al. Deep blue retrievals of Asian aerosol properties during ACE-asia[J].IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(11): 3180-3195.
- [8] Chen Q X, Han X L, Gu Y, et al. Evaluation of MODIS, MISR, and VIIRS daily level-3 aerosol optical depth products over land [J]. Atmospheric Research, 2022, 265:105810.
- [9] Gao L, Chen L, Li C, et al. Evaluation and possible uncertainty source analysis of JAXA Himawari-8 aerosol optical depth product over China[J].Atmospheric Research, 2021, 248:105248.

- [10] 胡秀清,卢乃锰,邱 红.FY-1C/1D 全球海上气溶胶业务反演 算法研究[J].海洋学报,2006,28(2):56-65.
  Hu X Q,Lu N M,Qiu H.Development of aerosol retrieval algorithm over global ocean using FY-1C/1D data[J].Acta Oceanologica Sinica,2006,28(2):56-65.
- [11]任通,高玲,李成才,等.利用风云 2C 静止卫星可见光资料 反演气溶胶光学厚度[J].北京大学学报(自然科学版),2011, 47(4):636-646.
   Ren T,Gao L,Li C C, et al. Retrieval of aerosol optical depth from

Fengyun-2C geostationary satellite observation: Theory and implementation[J]. Acta Scientiarum Naturalium Universitatis Pekinensis,2011,47(4):636-646.

- [12] 刘智超,王健,张鹏,等.风云三号卫星气溶胶光学厚度产品的适用性验证[J].环境监测管理与技术,2023,35(2):18-22.
  Liu Z C, Wang J, Zhang P, et al. Verification of aerosol optical depth observed by FY-3 satellite [J]. The Administration and Technique of Environmental Monitoring,2023,35(2):18-22.
- [13] Saide P E, Kim J, Song C H, et al. Assimilation of next generation geostationary aerosol optical depth retrievals to improve air quality simulations[J].Geophysical Research Letters, 2014, 41 (24):9188– 9196.
- [14] 许梦婕.FY-4A 气象卫星对京津冀地区 AOD 的反演方法研究
  [D].南京:南京信息工程大学,2021.
  Xu M J.Study on inversion method of AOD in Beijing-Tianjin-Hebei region by FY-4A meteorological satellite[D].Nanjing:Nanjing University of Information Science & Technology,2021.
- [15] Jiang X, Xue Y, Jin C, et al. A simple band ratio library (BRL) algorithm for retrieval of hourly aerosol optical depth using FY-4A AGRI geostationary satellite data [J]. Remote Sensing, 2022, 14 (19):4861.
- [16] 陈兴峰,郑逢杰,郭 丁,等.气溶胶定量遥感的机器学习方法综述[J].遥感学报,2021,25(11):2220-2233.
  Chen X F, Zheng F J, Guo D, et al. Review of machine learning methods for aerosol quantitative remote sensing[J]. National Remote Sensing Bulletin, 2021,25(11):2220-2233.
- [17] Ristovski K, Vucetic S, Obradovic Z. Uncertainty analysis of neuralnetwork-based aerosol retrieval [J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(2):409-414.
- [18] Levy R C, Remer L A, Kleidman R G, et al. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land[J]. Atmospheric Chemistry and Physics, 2010, 10 (21): 10399 – 10420.
- [19] Ding H, Zhao L, Liu S, et al. FY-4A/AGRI aerosol optical depth retrieval capability test and validation based on NNAeroG[J].Remote Sensing, 2022, 14(21):5591.
- [20] Yuan Q, Shen H, Li T, et al. Deep learning in environmental remote sensing: Achievements and challenges [J]. Remote Sensing of Environment, 2020, 241:111716.
- [21] Eck T F, Holben B N, Reid J S, et al. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols[J].Journal of Geophysical Research: Atmospheres, 1999, 104 (D24):31333-31349.
- [22] Sun Z, Zhang B, Yao Y. Improving the estimation of weighted mean temperature in China using machine learning methods [J]. Remote Sensing, 2021, 13(5):1016.

· 220 ·

- [23] 高阿芳.基于改进随机森林的耕深预测模型研究[D].长春:长春工业大学,2022.
   Gao A F.Research on prediction model of tillage depth based on an improved random forest[D].Changchun; Changchun University of Technology,2022.
- [24] Levy R C, Mattoo S, Munchak L A, et al. The Collection 6 MODIS aerosol products over land and ocean [J]. Atmospheric Measurement Techniques, 2013, 6(11):2989-3034.
- [25] Levy R C, Munchak L A, Mattoo S, et al. Towards a long-term global aerosol optical depth record: Applying a consistent aerosol retrieval algorithm to MODIS and VIIRS-observed reflectance [J]. Atmospheric Measurement Techniques, 2015, 8(10):4083-4110.
- [26] Sawyer V, Levy R C, Mattoo S, et al. Continuing the MODIS dark target aerosol time series with VIIRS[J].Remote Sensing, 2020, 12 (2):308.
- [27] Chen Y, Fan M, Li M, et al. Himawari-8/AHI aerosol optical depth detection based on machine learning algorithm [J]. Remote Sensing, 2022, 14(13): 2967.
- [28] 陈岚琪.基于 GWR 的郑州市域林地景观格局、气象与气溶胶

光学厚度的空间关系研究[D].郑州:河南农业大学,2020. Chen L Q.Spatial relationship among forest landscape pattern, meteorology and aerosol based on GWR model in Zhengzhou City[D]. Zhengzhou:Henan Agricultural University,2020.

- [29] 杨 阳,赵 娜,岳天祥.1980—2018 年中国极端高温事件时空格局演变特征[J].地理科学,2022,42(3):536-547.
  Yang Y,Zhao N, Yue T X.Spatio-temporal variations of extreme high temperature event in China from 1980 to 2018[J].Scientia Geographica Sinica,2022,42(3):536-547.
- [30] Xue R, Ai B, Lin Y, et al.Spatial and temporal distribution of aerosol optical depth and its relationship with urbanization in Shandong Province[J].Atmosphere, 2019, 10(3):110.
- [31] Zhao C, Wang Y, Shi X, et al. Estimating the contribution of local primary emissions to particulate pollution using high-density station observations [J]. Journal of Geophysical Research: Atmospheres, 2019, 124(3):1648-1661.
- [32] Jin Q, Wei J, Yang Z L, et al. Consistent response of Indian summer monsoon to Middle East dust in observations and simulations [J]. Atmospheric Chemistry and Physics, 2015, 15(17):9897-9915.

# Machine learning-based inversion of aerosol optical depth inversion from FY-4A data

# CHEN Xin, SHI Guoping

(School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: This study aims to develop a machine learning algorithm based on the characteristics of AGRI data to generate an aerosol dataset with a high spatiotemporal resolution. Using aerosol data from 67 aerosol robotic network (AERONET) sites in China and its surrounding areas in 2021, this study selected data of factors such as apparent reflectance, observation angles, elevation, and MODIS surface reflectance acquired from FY - 4A advanced geostationary radiation imager (AGRI) - a new generation geostationary meteorological satellite of China. Then, this study performed the inversion of aerosol optical depth (AOD) using four machine learning methods-random forest (RF), gradient boosting regression tree (GBRT), extreme gradient boosting (XGBoost), and back propagation neural network (BPNN). Using the optimal model determined based on evaluation metrics, this study achieved the AOD inversion with a spatial resolution of 4 km  $\times$  4 km based on FY-4A data. Then, this study compared the inversion results with MODIS aerosol products of roughly the same periods. The results indicate that the AOD inversion models based on the four machine learning algorithms yielded correlation coefficients (R) exceeding 0. 90, mean absolute errors (MAEs) of less than 0.09, and root mean square errors (RMSE) below 0.14. This indicates that it is feasible to conduct AOD inversion using machine learning-based models. The GBRT-based model exhibited the highest inversion accuracy among the four methods, with a correlation coefficient of 0.82, MAE of 0.16, and RMSE of 0.25, as indicated by the verification results. Additionally, 47% of the inversion results fell within the expected error ranges, indicating that the results of AOD inversion from FY-4A data using the GBRTbased model were generally consistent with observed values. The comparison between the GBRT model-derived AOD inversion results and the results of MODIS aerosol products shows that the former exhibited high consistency with the latter in terms of spatial distribution, with 83.57% of grid deviations falling within the range from -1.0 to 0 and the former slightly higher than the latter.

Keywords: machine learning; AOD; FY-4A/AGRI; AERONET; MODIS AOD