

# 基于多特征参数的 OMI 遥感产品气溶胶分类研究

——以广东省为例

郑仰成<sup>1,2,3</sup>, 黎丽莉<sup>4\*</sup>, 王云鹏<sup>1,2</sup>

(1. 中国科学院广州地球化学研究所有机地球化学国家重点实验室, 广州 510640;

2. 中国科学院深地科学卓越创新中心, 广州 510640; 3. 中国科学院大学, 北京 100049;

4. 仲恺农业工程学院资源与环境学院, 广州 510225)

**摘要:**为了解决气溶胶分类精度低和特征参数冗杂的问题,基于 OMI(Ozone Monitoring Instrument)遥感产品的气溶胶特征参数,利用随机森林算法,将广东省 2014 年的气溶胶类型划分为沙尘型气溶胶(Desert Dust, DST)、生物质燃烧型含碳气溶胶(Carbonaceous Aerosols Associated with Biomass Burning, CRB)和硫酸盐型城镇-工业气溶胶(Sulfate-based Urban-industrial Aerosols, SLF)3 种类型.并统计分析随机森林以及特征参数的重要性,将分类结果的空间分布与 OMI 气溶胶类型产品的空间分布进行对比.结果表明:(1)随机森林算法仅需少量训练样本点即可达到 97% 以上的总体分类精度.(2)通过计算不同气溶胶特征参数在随机森林分类过程中的重要性高低,得到重要性排名前六的特征参数依次为  $\alpha$  指数、UVAI、 $RI_{388}$ 、 $RI_{354}$ 、 $SSA_{500}$ 、 $AAOD_{500}$ ,表明在分类过程中,气溶胶粒径分布和吸收能力起到了最关键的作用.(3)3 种气溶胶类型的空间分布显示,SLF 型气溶胶为广东省最主要的气溶胶类型;DST 型和 CRB 型气溶胶在珠三角地区占比最高,在粤东、粤北地区的占比最低.

**关键词:**气溶胶分类; 随机森林; 臭氧层监测仪; 特征参数;  $\alpha$  指数; UVAI

中图分类号:X87 文献标志码:A 文章编号:1000-5463(2021)04-0068-08

## The Multi-feature Parameter Classification of Aerosol Based on OMI Remote Sensing Data: A Case Study in Guangdong Province

ZHENG Yangcheng<sup>1,2,3</sup>, LI Lili<sup>4\*</sup>, WANG Yunpeng<sup>1,2</sup>

(1. Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China;

2. Center for Excellence in Deep Earth Science, Chinese Academy of Sciences, Guangzhou 510640, China;

3. College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 510640, China;

4. College of Environmental Science and Engineering, Zhongkai University of Agricultural and Engineering, Guangzhou 510225, China)

**Abstract:** In order to solve the problems of low precision and redundant feature parameters in the process of aerosol classification, several significant aerosol feature parameters were extracted from the OMI (Ozone Monitoring Instrument) remote sensing products and Random Forest (RF) algorithm was used for aerosol type classification and verification. Aerosols of Guangdong Province in 2014 were divided into three types: desert dust (DST), carbonaceous aerosols associated with biomass burning (CRB) and sulfate-based urban-industrial aerosols (SLF). The Random Forest classification results and the importance of feature parameters were analyzed. The spatial distribution of classification results were compared to that of OMI aerosol type products. The following results are obtained. First, with the RF algorithm, a total precision of over 97% can be reached with a few training samples. Second, calculating the importance of different aerosol feature parameters in RF shows that the most important feature parameters are angstrom exponent, UVAI,  $RI_{388}$ ,  $RI_{354}$ ,  $SSA_{500}$  and  $AAOD_{500}$  in turn, indicating that size distribution and absorption ability of aerosols play the key roles. Third, the spatial distribution of three aerosol types shows that sulfate-based urban-industrial aerosols are dominant in the Pearl River Delta. Proportion of biomass burning and desert dust aerosols are highest in the Pearl River Delta and lowest in the eastern and northern Guangdong.

**Keywords:** aerosol classification; random forest; OMI; feature parameter; angstrom exponent; UVAI

大气中的气溶胶粒子通过散射、吸收太阳辐射和形成云凝结核等方式来影响全球气候变化<sup>[1]</sup>. 不同类型气溶胶之间的光学属性、成分以及粒径分布有所差异,因此,不同类型的气溶胶也会给气候带来不同程度的影响<sup>[2]</sup>. 沙尘型气溶胶的粒径分布较大,有较强的散射能力,会降低地表的能见度<sup>[3]</sup>. 生物质燃烧型含碳气溶胶的粒径分布较小,但由于吸收性成分含量较高,此类气溶胶表现出较强的辐射吸收能力;大量的生物质燃烧型含碳气溶胶会影响云的微物理性质,甚至影响生物地球化学循环<sup>[4]</sup>. 因此,对不同的气溶胶类型进行有效的区分,有助于进行气溶胶源解析、改善气溶胶光学模型、减少卫星观测气溶胶的不确定性<sup>[5]</sup>.

传统的监测和分析气溶胶类型的方法主要基于地面的监测站点,这种监测分析方法受限于监测站点的数量以及空间分布,因此,地面站点的监测数据无法反映真实的气溶胶空间分布情况<sup>[6]</sup>. 遥感技术具有监测范围大、覆盖时间长的特点,在气溶胶的监测分析中起到了不可或缺的作用,利用卫星遥感数据可以进行广覆盖、长时间序列的气溶胶监测<sup>[7-9]</sup>. 利用遥感数据进行气溶胶分类,其原理是提取描述气溶胶的消光、吸收、散射、折射和粒径分布等各种气溶胶特征参数. 不同种类的气溶胶的各项特征参数互有差别,因此,可以通过特征参数区分不同的气溶胶类型<sup>[4,10-11]</sup>.

目前已有许多大气遥感卫星发布了气溶胶相关的产品,为气溶胶的监测和分析发挥了重要的作用. OMI AuraAER 是美国国家航空航天局(National Aeronautics and Space Administration, NASA)公布的一组近紫外型气溶胶产品,该产品由搭载在 Aura 卫星上的 OMI(Ozone Monitoring Instrument)传感器观测数据反演得到,包含多项气溶胶特征参数以及气溶胶产品类型<sup>[12]</sup>. 对 EOS-Aqua 卫星上的一氧化碳指数(Carbon Monoxide Index, COI)和 OMI 上的近紫外型气溶胶指数(Ultraviolet Aerosol Index, UVAI)进行阈值分析,可得到 OMI 的 3 种气溶胶产品类型:沙尘型气溶胶(Desert Dust, DST)、生物质燃烧型含碳气溶胶(Carbonaceous Aerosols Associated with Biomass Burning, CRB)和硫酸盐型城镇-工业气溶胶(Sulfate-based Urban-industrial Aerosols, SLF). 这 3 类气溶胶产品类型已得到广泛运用,对气溶胶空间

特征研究、源解析及气溶胶光学模型的改善起到了重要的作用<sup>[13-15]</sup>. 但是,COI 数据和 UVAI 数据来自 2 颗不同的卫星,时间分辨率与空间分辨率均存在差异,无法满足数据融合的需求,部分像元点无法反演得到气溶胶类型的数据. 因此,本研究以广东省为研究区,在不使用 COI 辅助数据条件下进行气溶胶类型的反演,从而提高气溶胶类型产品的空间覆盖度,拟为广东省的气溶胶源解析和排放源控制提供科学参考.

## 1 研究数据与方法

### 1.1 研究数据

以广东省(20°13' N ~ 25°31' N, 109°39' E ~ 117°19' E)为研究区域,采用 OMI 的气溶胶产品进行研究. NASA 的近紫外型气溶胶产品由不同年份获取的地表反射率及其反演得到的气溶胶产品构成,其中 OMI AuraAER 为 2004 年 8 月至今的气溶胶产品. 该产品的时间分辨率为 98.8 min、空间分辨率为 13 km×24 km,2 个用于获取地表反射率的波段分别为 354、388 nm,并由这 2 个波段反演得到一个 500 nm 波段的产品. 该卫星为极地轨道卫星,观测范围可以覆盖全球所有区域,本研究仅选用覆盖研究区域内的像元点进行研究.

在 OMI AuraAER 产品中,有多个用于反映气溶胶不同物理属性的气溶胶特征参数,其中可被用于气溶胶分类的气溶胶特征参数包括:

(1) 气溶胶吸收光学厚度(Aerosol Absorption Optical Depth, AAOD):描述气溶胶对太阳辐射的消光作用中吸收的部分,取值范围为 0~1.

(2) 气溶胶光学厚度(Aerosol Optical Depth, AOD):描述气溶胶对太阳辐射的总消光能力,取值范围为 0~6.

(3) 单向散射反照率(Single Scattering Albedo, SSA):描述气溶胶的吸收和散射的相对大小,SSA 值越大,则说明该气溶胶的散射能力强于吸收能力,取值范围为 0~1.

(4) 复折射指数(Refraction Index, RI):由实部和虚部组成,其中实部描述气溶胶的散射能力,虚部描述气溶胶的吸收能力. OMI AuraAER 产品仅提供了 RI 的虚部,取值范围为 0~1.

(5) 近紫外型气溶胶指数 (Ultraviolet Aerosol Index, UVAI): 反映吸气溶胶吸收能力的强弱. 吸收性气溶胶的 UVAI 呈现正值, 且吸收性越强 UVAI 值越高, 而非吸收性气溶胶的 UVAI 呈现负值. 取值范围为  $-2 \sim 3$ .

(6)  $\alpha$  指数 (Angstrom Exponent): 通过 2 个波段 (388 nm 和 500 nm) 的 AOD 产品计算得到, 反映气溶胶粒径分布的大小. 气溶胶粒径分布越大, 则  $\alpha$  值越小, 反之则越大. 取值范围为  $0.5 \sim 2.0$ .

(7) 经纬度: 每个像元点的经度和纬度坐标, 数值范围为广东省的经纬度范围.

本文使用上述所有气溶胶特征参数作为训练数据, 包括 AAOD ( $\lambda$  为 354、388、500 nm)、AOD ( $\lambda$  为 354、388、500 nm)、SSA ( $\lambda$  为 354、388、500 nm)、RI ( $\lambda$  为 354、388 nm)、UVAI、经纬度以及  $\alpha$  指数共计 15 个特征参数; 以 OMIAuraAER 自带的气溶胶类型产品作为标签数据, 利用随机森林算法对广东省气溶胶类型进行监督分类. ZHENG 等<sup>[16]</sup> 曾使用 K-means 聚类算法, 结合 AOD、UVAI 和  $\alpha$  指数 3 种气溶胶特征参数对 2010 年广东省气溶胶类型进行非监督分类. 为形成长时间序列研究, 本研究的时间段选取 2014 年, 探讨不同的分类方法及特征参数选取对气溶胶分类的影响. 选取数据集中所有产品质量合格, 各特征参数没有异常值的像元点共计 53 882 个, 其中: SLF 型气溶胶占比最高 (84.3%), 共有 45 404 个像元点; 其次是 DST 型气溶胶占比 (12.3%), 有 6 629 个像元点; CRB 型气溶胶占比最低 (3.4%), 有 1 849 个像元点.

## 1.2 研究方法

在使用 OMI 气溶胶产品的特征参数进行气溶胶分类的过程中存在 2 个问题: (1) 如何在多种特征参数中挑选出最适合分类的特征参数及其组合, 使得分类结果的精确度和计算性能满足研究需求. (2) 已有研究使用的分类模型大多是基于阈值分类法, 即在特征参数上确定一个阈值, 以区别不同的气溶胶类型. 然而, 这个阈值大多由历史经验数据给出, 而且不同地区的气溶胶的物理特征有所差异, 在一个地区所确定的阈值并不能广泛地运用到其他区域进行分类.

本文利用随机森林算法进行气溶胶分类, 探讨

适用于气溶胶分类的特征参数及其组合. 随机森林算法<sup>[17]</sup> 是一种分类速度快、分类准确率高的监督分类算法, 已用于解决土地覆盖类型分类、植被覆盖度估算、农作物识别和用水总量影响因素解析等问题, 并取得了良好的效果<sup>[18-24]</sup>. 在分类过程中, 利用随机森林算法可以得到不同特征参数的重要性, 可根据重要性的高低对气溶胶的各项特征参数进行排序, 从而挑选其中重要性最高的特征参数. 此外, 随机森林并不会显式地给出一个具体的分类阈值, 而是根据不同的训练数据得到不同的模型, 泛化性能要高于传统的阈值分类法.

随机森林由多棵决策树构成, 每棵决策树会随机选取样本数据及气溶胶特征参数, 得到一个预测的气溶胶类别, 然后由多棵决策树进行众数投票, 选取得票最多的预测类别作为随机森林的预测结果 (图 1A). 在决策树中, 用 Gini 指数来衡量数据集的不确定性, Gini 指数值越大, 则表示数据集的不确定性越大. 决策树的分类过程即是通过不断分割训练样本, 从而降低样本数据集 Gini 指数的过程. 对于整个样本数据集  $D$ , 其 Gini 指数为:

$$\text{Gini}(D) = 1 - \sum_{k=1}^3 \left( \frac{|C_k|}{|D|} \right)^2, \quad (1)$$

其中,  $k=1, 2, 3$ , 代表有 3 种气溶胶类型 (DST、CRB 和 SLF);  $C_k$  是指每个类的样本数.

根据特征参数  $A$ , 将数据集  $D$  分割成 2 个子数据集 ( $D_1$  和  $D_2$ ), 则在特征参数  $A$  的条件下, 数据集  $D$  的 Gini 指数为:

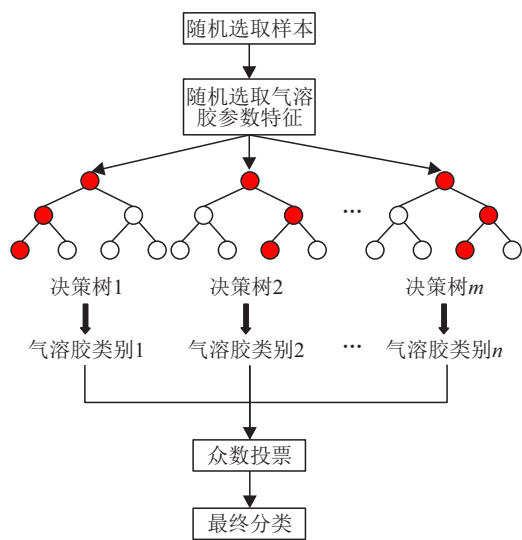
$$\text{Gini}(D, A) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2), \quad (2)$$

即数据集  $D$  在特征参数  $A$  条件下的 Gini 指数可以用 2 个子数据集的 Gini 指数来表示. 通过寻找最优的阈值, 使数据集的 Gini 指数在分割成子数据集后降至最低, 在子数据集中尽量只包含属于同一个类的样本点. 利用数据集  $D$  和子数据集  $D_1$ 、 $D_2$  在特征参数  $A$  条件下的 Gini 指数之差来衡量特征参数  $A$  的重要性.

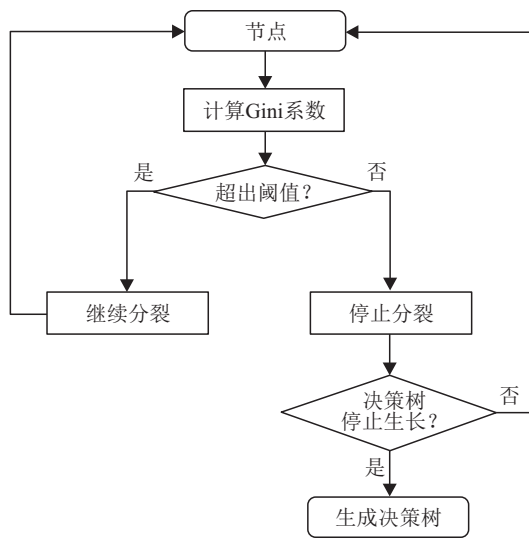
在每棵决策树中, 每个节点代表一个样本数据集. 计算每个节点的 Gini 指数, 如果 Gini 指数高于阈值, 则代表该节点的不确定性过高, 需要继续分裂以降低不确定性; 如果 Gini 指数低于阈值, 则代表该节点是一个稳定的分类结果, 可以停止分裂, 形成

一个叶子结点. 如果所有叶子结点不再分裂, 即决策树不再成长, 则输出该决策树, 再继续随机选取不

同的样本点和气溶胶特征参数构建新的决策树, 直到决策树的数量达到要求(图 1B).



(A)随机森林分类过程



(B)单棵决策树分类过程

图 1 气溶胶分类流程图

Figure 1 The flow chart of aerosol type classification

在随机森林算法中, 使用特征参数重要性来衡量每个特征参数在分类过程中作出的贡献大小. 在决策树中, 当 1 个节点分裂为 2 个子节点后, 数据集被分为 2 个确定度更高的子集, 衡量不确定性的 Gini 指数也相应减小. 如果一个特征参数能够降低的不确定性越多, 则其分类效果越好, 重要性也越高. 将每个特征参数的重要性进行归一化, 使其重要性之和为 1.

在随机森林的参数设置中, 决策树的数量为 100 棵. 由于标签数据仅有 3 种类别 (DST、CRB、SLF), 而用于分类的特征数量有 15 个, 为了防止每棵树的分裂深度过大, 将最大叶子结点数量设置为 6 个, 即每棵决策树出现 6 个叶子结点后将停止分裂, 以节省分类时间. 在研究过程中, 依次选取不同数量的样本点进行训练, 观察随机森林算法的分类精度与样本点数量的关系, 并使用所有的样本点进行精度的验证.

## 2 结果与讨论

基于 2014 年广东省气溶胶类型的分类结果, 本文讨论了随机森林算法的分类精度, 提取分类过程中重要的气溶胶特征参数, 并对气溶胶类型的空间分布特征进行了分析.

### 2.1 分类精度

分类精度是衡量算法优劣的重要指标. 计算分类正确的训练样本数量在总体训练样本数量中的百分比, 作为总体的分类精度. 同时, 分别计算 DST、CRB、SLF 型气溶胶分类正确的训练样本数量在各自类别的训练样本数量中的百分比, 作为各个类别的分类精度.

由分类精度结果(图 2)可知:(1)总体的分类精度可达 97% 以上, 且随着投入训练的像元数量增加而不断提高.(2)初始训练样本数量较少时, CRB 型气溶胶的分类精度较低(仅为 35% 左右), 但由于 CRB 样本在总体样本中所占比例较低, 因此, 总体的分类精度和其他 2 类气溶胶类型(DST, SLF)的分类精度均可达到 60%~70%.(3)随着训练样本数量的增加, 总体的分类精度与各个分类别的分类精度迅速增大. 训练样本数量达到 50 个时, DST、SLF 型气溶胶和总体的分类精度均可达到 97% 以上, 而 CRB 型气溶胶的分类精度有点波动, 但也逐渐趋于稳定. 由此可知, 使用随机森林算法进行气溶胶分类, 可以满足在不使用 COI 辅助数据的条件下, 单独用 OMI 自身的气溶胶特征参数完成气溶胶的高精度分类; 同时, 在实际应用中, 受到卫星过境时间以及云层覆盖度的影响, 单景遥感影像中可用于训练的高质量像元数量并不多, 而随机森林算法的高效性使其可以在训练像元数量稀疏的特殊情况下建立分类模型.

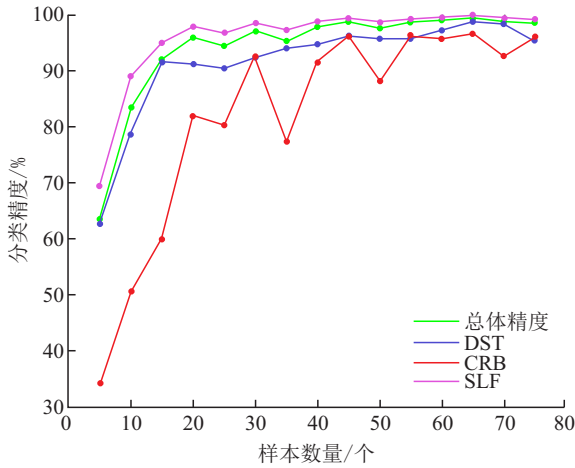


图 2 气溶胶分类精度与训练样本数量关系

Figure 2 The relationship between the training sample size and the classification precision

## 2.2 特征参数选取

由特征参数重要性结果(表 1)可知:(1)在所有特征参数中, $\alpha$  指数对气溶胶分类影响最大,其重要性达到 0.245,说明气溶胶的粒径分布是用于判断气溶胶类型的重要指标。(2)UVAI 指数的重要性达到 0.218,在各项气溶胶特征参数中排名第二,符合预期结果. 究其原因为:在反演 OMI 气溶胶产品的阈值分类法中,使用了 UVAI、AOD、SSA 以及 COI 指数,其中 UVAI 占主导因素。(3)在其余特征参数中,重要性由高到低依次为 RI、SSA、AAOD、AOD,不同波段的 AAOD、AOD、SSA 和 RI 的重要性差别不大. 因此,在使用气溶胶分类算法中,使用哪个波段的产品对于分类精度没有太大影响。(4)经纬度对气溶胶分类的重要性接近 0,意味着经纬度在分类过程中无法降低数据集的不确定性,是无效的分类特征参数.

由分类价值最高的 6 个特征参数( $\alpha$  指数、UVAI、 $RI_{354}$ 、 $RI_{388}$ 、 $SSA_{500}$  和  $AAOD_{500}$ ) 在 3 种气溶胶类型中的均值、标准差以及最大、最小值(表 2)可知:(1)在反映气溶胶粒径分布的  $\alpha$  指数上,SLF 型气溶胶的均值最大,说明其粒径分布最小;而 DST 型气溶胶的均值较小,说明其粒径分布最大。(2)在反映气溶胶吸收能力的 UVAI 指数上,DST、CRB 型气溶胶的均值较大,说明这 2 种气溶胶的吸收能力强;SLF 型气溶胶的均值最小,说明 SLF 型气溶胶的吸收能力较弱。(3)CRB 型气溶胶的  $\alpha$  指数均值与 SLF 型气溶胶的相近、UVAI 均值与 DST 型气溶胶的相近,单独考虑一种特征参数无法较好地识别 CRB 型气溶胶. 此时可结合  $\alpha$  指数和 UVAI 指数进行判别:CRB 型气溶胶在这 2 种指数上均具有较高的均值,可以凭此区分 CRB 型气溶胶与另外 2 种气溶胶类型.

表 1 各项气溶胶特征参数的重要性

Table 1 The importance of each aerosol feature parameter

特征参数	重要性
AAOD $\lambda = 354 \text{ nm}$	0.034
$\lambda = 388 \text{ nm}$	0.019
$\lambda = 500 \text{ nm}$	0.045
AOD $\lambda = 354 \text{ nm}$	0.008
$\lambda = 388 \text{ nm}$	0.009
$\lambda = 500 \text{ nm}$	0.008
SSA $\lambda = 354 \text{ nm}$	0.038
$\lambda = 388 \text{ nm}$	0.020
$\lambda = 500 \text{ nm}$	0.079
RI $\lambda = 354 \text{ nm}$	0.141
$\lambda = 388 \text{ nm}$	0.128
经度	0.004
纬度	0.004
UVAI	0.218
$\alpha$ 指数	0.245

表 2 3 类气溶胶特征参数的均值、标准差、最大值及最小值  
Table 2 The mean value, standard deviation, and maximum and minimum values of three aerosol feature parameter

气溶胶类型	特征参数	均值	标准差	最大值	最小值
DST	$\alpha$ 指数	0.605	0.000	0.604	0.602
	UVAI	1.233	0.416	3.862	0.800
	$RI_{388}$	0.007	0.004	0.028	0.003
	$RI_{354}$	0.019	0.010	0.071	0.007
	$SSA_{500}$	0.950	0.026	0.982	0.829
	$AAOD_{500}$	0.039	0.018	0.155	0.011
CRB	$\alpha$ 指数	1.614	0.059	1.740	1.544
	UVAI	1.168	0.460	5.604	0.801
	$RI_{388}$	0.001	0.001	0.016	0.000
	$RI_{354}$	0.002	0.002	0.022	0.000
	$SSA_{500}$	0.985	0.010	0.997	0.869
	$AAOD_{500}$	0.016	0.007	0.082	0.005
SLF	$\alpha$ 指数	1.853	0.013	1.872	1.779
	UVAI	0.106	0.385	0.800	-2.186
	$RI_{388}$	0.004	0.003	0.019	0.000
	$RI_{354}$	0.004	0.003	0.019	0.000
	$SSA_{500}$	0.964	0.021	1.000	0.853
	$AAOD_{500}$	0.026	0.014	0.082	0.000

## 2.3 广东省气溶胶类型特征讨论

在使用随机森林法进行气溶胶分类后,可以统计广东省总体及各个城市中各类型气溶胶的占比. 由图 3 和表 3 可知:(1)在广东省中,SLF 型气溶胶占比最高,达到 84.5%;其次是 DST 型气溶胶占比,达到 12.5%;CRB 型气溶胶占比最低(3.0%). (2)SLF 型气溶胶在广东省内占比最高. SLF 型气溶胶主要来自工业生产和机动车排放,因此,可反映广东

省的工业化程度较高,机动车数量较多。(3)DST 型气溶胶在珠三角地区的占比最高,其次是在粤西地区,在粤东地区的占比最低。DST 型气溶胶主要来自道路扬尘与土木工程建设,反映出珠三角和粤西地区的交通设施较为发达,在建城市基础设施较多。(4)CRB 型气溶胶在珠三角及粤东汕头、潮州等城市占比较高。珠三角地区的 CRB 型气溶胶主要来自大量的汽车尾气排放,而粤东汕头和潮州地区的 CRB 型气溶胶则可能来自秸秆的燃烧。

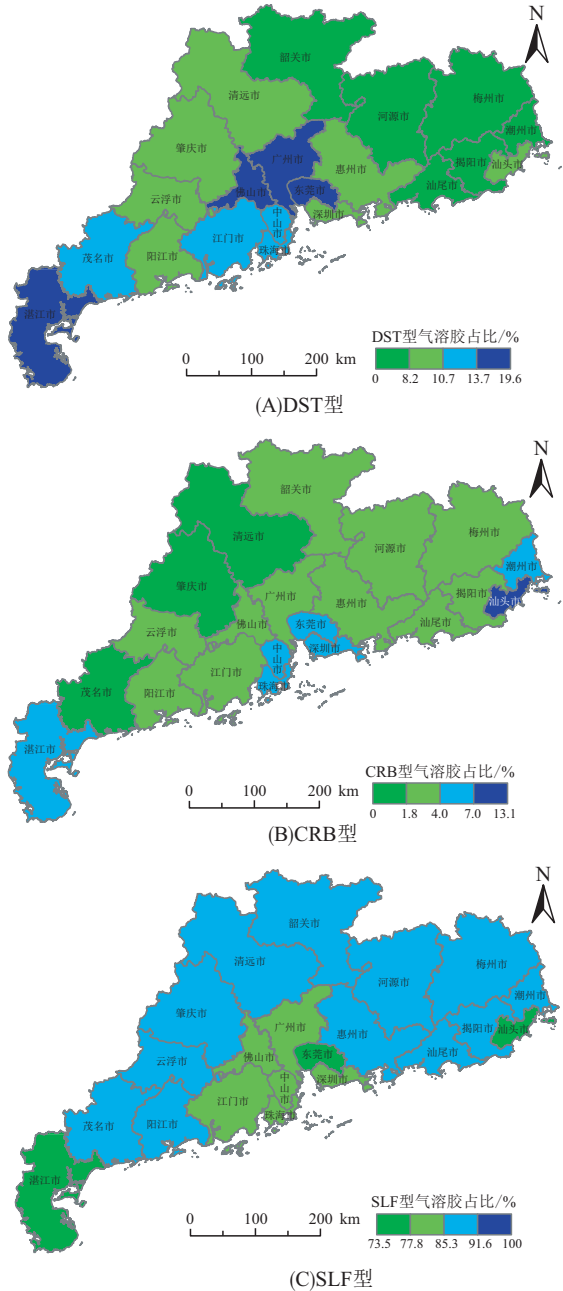


图 3 广东省的气溶胶类型占比的空间分布(随机森林分类结果)

Figure 3 The spatial distribution of aerosol types proportion in Guangdong (results of RF classification)

表 3 广东省各城市 3 种气溶胶类型占比

Table 3 The proportion of three aerosol types of each city in Guangdong %

城市	OMI 产品标签数据			随机森林分类结果		
	DST	CRB	SLF	DST	CRB	SLF
潮州市	6.6	2.5	90.9	6.6	4.1	89.3
东莞市	23.9	1.2	74.9	19.6	4.3	76.1
佛山市	16.4	1.3	82.3	19.6	2.2	78.2
广州市	17.1	1.8	81.1	17.1	2.1	80.8
河源市	7.8	1.2	9.1	7.8	1.9	90.3
惠州市	9.6	1.0	89.4	9.6	2.5	87.9
江门市	13.8	1.6	84.6	13.7	1.9	84.4
揭阳市	7.3	1.7	91.0	7.3	1.9	90.8
茂名市	12.8	0.7	86.5	12.8	0.7	86.5
梅州市	5.9	1.7	92.4	5.9	3.0	91.1
清远市	10.8	0.5	88.7	10.6	0.5	88.9
汕头市	8.6	8.6	82.8	10.6	11.6	77.8
汕尾市	6.9	4.0	89.1	6.9	3.9	89.2
韶关市	9.6	1.8	88.6	8.1	1.9	90.0
深圳市	10.1	2.9	87.0	10.1	5.0	84.9
阳江市	7.7	2.1	90.2	8.2	2.1	89.7
云浮市	5.3	1.1	93.6	8.2	1.9	89.9
湛江市	12.7	3.2	84.1	16.7	6.9	76.4
肇庆市	10.5	0.9	88.6	10.5	0.9	88.6
中山市	8.4	3.1	88.5	10.8	4.1	85.1
珠海市	8.7	3.3	88.0	10.8	4.1	85.1

由随机森林分类结果与 OMI 气溶胶类型产品标签数据的空间分布(图 3、图 4)可知:(1)两者的空间分布呈现高度的相似性。(2)各类型气溶胶占比排名相同,从高至低依次均为 SLF、DST、CRB 型气溶胶。(3)结合表 3 可知各种气溶胶类型的占比有所差异,CRB 型气溶胶占比升高,DST、SLF 型气溶胶占比降低。

本文所得 2014 年的分类结果与 2010 年的分类结果<sup>[16]</sup>相符,各气溶胶类型占比没有太大变化,说明广东省内气溶胶类型在 2010—2014 年间没有较大的变化。与 2010 年使用的非监督分类 K-means 聚类法相比,使用 OMI 气溶胶类型产品作为标签数据进行随机森林监督分类提高了分类结果的可靠性。本文提取了 6 项最重要的气溶胶特征参数,与 2010 年使用的 3 项特征参数相比, $\alpha$  指数和 UVAI 仍是重要的分类指标,但是 AOD 的重要性下降。

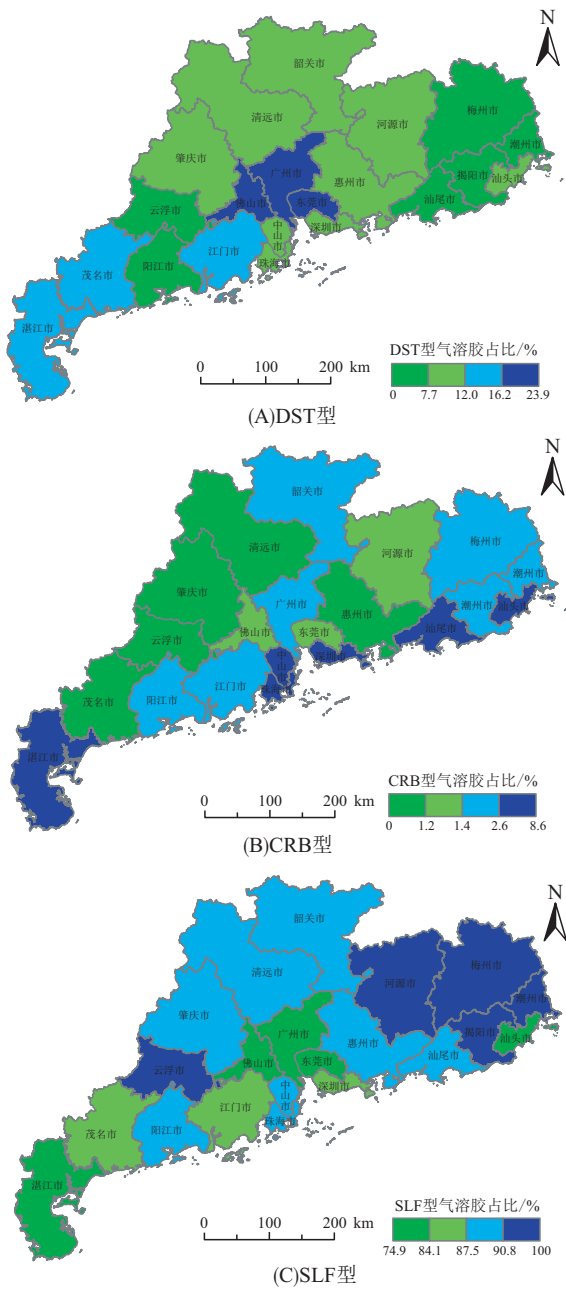


图 4 广东省的气溶胶类型占比的空间分布 (OMI 产品标签数据)

Figure 4 The spatial distribution of aerosol type proportion in Guangdong (label data from OMI product)

### 3 结论与展望

本研究基于 OMI 卫星遥感产品提取的特征参数,使用随机森林法对广东省 2014 年气溶胶进行分类并做验证. 根据随机森林法的分类结果,讨论了分类的精度、各项气溶胶特征参数的选取以及分类结果的空间分布,主要结论如下:

(1)使用随机森林法进行气溶胶分类,分类精

度可稳定在 97%以上,仅需少量样本点即可达到高精度的分类. 样本数量达到 50 个时,总体的分类精度可达 97.6%,其中 DST、CRB、SLF 型气溶胶的分类精度分别为 95.6%、88.1%、98.6%. 说明在不使用 COI 指数的条件下,通过随机森林算法可以用 OMI 自身的气溶胶特征参数产品完成较高精度的气溶胶分类.

(2)本研究使用了 15 个特征参数进行分类,根据特征参数重要性高低,最重要的 6 个分类特征参数依次为  $\alpha$  指数、UVAI、 $RI_{388}$ 、 $RI_{354}$ 、 $SSA_{500}$ 、 $AAOD_{500}$ ,说明气溶胶的粒径分布特征和光学吸收特征在分类中起到了最关键的作用.

(3)2014 年广东省内各城市各类型气溶胶的占比结果显示:SLF 型气溶胶为广东省最主要的气溶胶类型;DST 型气溶胶在珠三角地区的占比最高,其次是在粤西地区,在粤东地区的占比最低;CRB 型气溶胶在珠三角地区和潮州、汕头、湛江等城市的占比较高,在其他地区的较低.

本文提出了一种基于 OMI 气溶胶特征参数和随机森林算法的气溶胶分类法,为传统气溶胶分类方法存在的精度低和特征参数冗余的问题提供了一种解决方法. 在今后的研究中,将以本文的研究结果为基础,进一步反演 2018 年广东省的气溶胶类型,讨论更长时间序列气溶胶类型的变化特征.

### 参考文献:

- [1] EALO M, ALASTUEY A, PEREZ N, et al. Impact of aerosol particle sources on optical properties in urban, regional and remote areas in the north-western Mediterranean [J]. *Atmospheric Chemistry and Physics*, 2018, 18: 1149–1169.
- [2] CHEN Q X, SHEN W X, YUAN Y, et al. Verification of aerosol classification methods through satellite and ground-based measurements over Harbin, Northeast China [J]. *Atmospheric Research*, 2019, 216: 167–175.
- [3] RENARD J B, DULAC F, DURAND P, et al. In situ measurements of desert dust particles above the western Mediterranean Sea with the balloon-borne Light Optical Aerosol Counter/sizer (LOAC) during the ChArMEx campaign of summer 2013 [J]. *Atmospheric Chemistry and Physics*, 2018, 18: 3677–3699.
- [4] CRUTZEN P J, ANDREAE M O. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles [J]. *Science*, 1990, 250: 1669–1678.
- [5] DUBOVIK O, HOLBEN B, ECK T F, et al. Variability of absorption and optical properties of key aerosol types observed in worldwide locations [J]. *Journal of the Atmospheric Sciences*, 2001, 59: 590–608.

- [6] HAMILL P, GIORDANO M, WARD C, et al. An AERONET-based aerosol classification using the Mahalanobis distance[J]. *Atmospheric Environment*, 2016, 140: 213–233.
- [7] KALAPUREDDY M C R, KASKAOUTIS D G, RAJ P E, et al. Identification of aerosol type over the Arabian Sea in the pre-monsoon season during the Integrated Campaign for Aerosols, Gases and Radiation Budget (ICARB) [J]. *Journal of Geophysical Research*, 2009, 114: D17203/1–12.
- [8] SREEKANTH V. On the classification and sub-classification of aerosol key types over south central peninsular India; MODIS-OMI algorithm[J]. *Science of the Total Environment*, 2014, 468/469: 1086–1092.
- [9] 白冰, 张强, 陈旭辉, 等. 中国西北干旱半干旱区气溶胶分类及特征[J]. *中国沙漠*, 2019, 39(5): 105–110.
- BAI B, ZHANG Q, CHEN X H, et al. Classification and characteristics of aerosols in arid and semi-arid areas of northwest China[J]. *Journal of Desert Research*, 2019, 39(5): 105–110.
- [10] KUMAR K R, KANG N, YIN Y. Classification of key aerosol types and their frequency distributions based on satellite remote sensing data at an industrially polluted city in the Yangtze River Delta, China[J]. *International Journal of Climatology*, 2018, 38: 320–336.
- [11] CHEN Q X, YUAN Y, SHUAI Y, et al. Graphical aerosol classification method using aerosol relative optical depth [J]. *Atmospheric Environment*, 2016, 135: 84–91.
- [12] TORRES O, TANSKANEN A, VEIHELMANN B, et al. Aerosols and surface UV products from Ozone Monitoring Instrument observations: an overview[J]. *Journal of Geophysical Research-Atmospheres*, 2007, 112: D24S47/1–14.
- [13] BUCHARD V, DA SILVA A M, COLARCO P R, et al. Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis[J]. *Atmospheric Chemistry and Physics*, 2015, 15: 5743–5760.
- [14] FIELD R D, VAN DER WERF G R, FANIN T, et al. Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Niño-induced drought [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113: 9204–9209.
- [15] CHEN J M, LI C L, RISTOVSKI Z, et al. A review of biomass burning: emissions and impacts on air quality, health and climate in China[J]. *Science of the Total Environment*, 2017, 579: 1000–1034.
- [16] ZHENG Y C, LI L L, WANG Y P. An aerosol type classification method based on remote sensing data in Guangdong, China[C]//The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Göttingen, Germany: Copernicus Publications, 2019: 239–243.
- [17] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, 45: 5–32.
- [18] 耿仁方, 付波霖, 蔡江涛, 等. 基于无人机影像和面向对象随机森林算法的岩溶湿地植被识别方法研究[J]. *地球信息科学学报*, 2019, 21(8): 1295–1306.
- GENG R F, FU B L, CAI J T, et al. Object-based karst wetland vegetation classification method using unmanned aerial vehicle images and random forest algorithm[J]. *Journal of Geo-information Science*, 2019, 21(8): 1295–1306.
- [19] 曹爽, 潘锁艳, 管海燕. 机载多光谱 LiDAR 的随机森林地物分类[J]. *测绘通报*, 2019, 11: 79–84.
- CAO S, PAN S Y, GUAN H Y. Random forest-based land-use classification using multispectral LiDAR data[J]. *Bulletin of Surveying and Mapping*, 2019, 11: 79–84.
- [20] 马玥, 姜琦刚, 孟治国, 等. 基于随机森林算法的农耕地土地利用分类研究[J]. *农业机械学报*, 2016, 47(1): 297–303.
- MA Y, JIANG Q G, MENG Z G, et al. Classification of land use in farming area based on random forest algorithm [J]. *Transactions of the Chinese Society of Agriculture Machinery*, 2016, 47(1): 297–303.
- [21] SU H, SHEN W, WANG J, et al. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests [J]. *Forest Ecosystems*, 2020, 7: 64/1–20.
- [22] HUNT D A, TABOR K, HEWSON J H, et al. Review of remote sensing methods to map coffee production systems [J]. *Remote Sensing*, 2020, 12: 2041/1–23.
- [23] 李宁, 汪雨娜. 基于随机森林回归算法的用水总量影响因素解析——以广东省为例[J]. *华南师范大学学报(自然科学版)*, 2021, 53(1): 78–84.
- LI N, WANG L N. An analysis of the factors in total water consumption based on random forest regression algorithm: a case study of Guangdong Province[J]. *Journal of South China Normal University(Natural Science Edition)*, 2021, 53(1): 78–84.
- [24] 刘望保, 谢智豪. 位置服务大数据下广州市土地利用类型模拟探讨[J]. *华南师范大学学报(自然科学版)*, 2019, 51(1): 75–83.
- LIU W B, XIE Z H. Inferring land use of Guangzhou from big data of location service[J]. *Journal of South China Normal University(Natural Science Edition)*, 2019, 51(1): 75–83.