

光学学报

一种基于朴素贝叶斯分类器的气溶胶类型识别模型

周妹¹, 常建华^{1,2*}, 陈思成¹, 孟园园¹, 戴腾飞^{1,2}

¹南京信息工程大学电子与信息工程学院, 江苏 南京 210044;

²南京信息工程大学江苏省大气环境与装备技术协同创新中心, 江苏 南京 210044

摘要 根据 AERONET SGP 站的气溶胶光学反演数据, 提出了一种基于朴素贝叶斯分类器的气溶胶分类模型, 以气溶胶单次散射反照率、复折射率指数等作为输入变量识别了该地区 5 种类型的气溶胶, 分析了不同类型气溶胶的光学特性。所提模型基于学习训练集样本的类别概率分布生成分类器模型, 进而预测出测试集样本的类别。在此基础上, 将所提模型应用于 SGP 站气溶胶类型的季节分布差异特性分析, 实验结果符合该地区的气候环境特征。为研究所提模型分类结果的准确性, 利用 AERONET 站点数据与高光谱分辨率激光雷达廓线数据匹配法建立了气溶胶分类阈值标准。结果表明, 相比传统的气溶胶分类算法, 所提模型获得的气溶胶分类结果与基于阈值标准确定的结果具有较高的一致性, 可为卫星等遥感设备的气溶胶反演提供地面数据支撑。

关键词 大气光学; 气溶胶分类; 机器学习; 朴素贝叶斯分类器; 气溶胶特性

中图分类号 P413

文献标志码 A

DOI: 10.3788/AOS202242.1801006

Aerosol Type Recognition Model Based on Naive Bayesian Classifier

Zhou Mei¹, Chang Jianhua^{1,2*}, Chen Sicheng¹, Meng Yuanyuan¹, Dai Tengfei^{1,2}

¹*School of Electronics & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;*

²*Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China*

Abstract Based on the aerosol optical inversion data from AERONET SGP station, an aerosol classification model based on a naive Bayesian classifier is proposed. The single scattering albedo and complex refractive index of aerosol are used as input variables to identify five types of aerosols in this region, and the optical properties of different types of aerosols are analyzed. The proposed model generates a classifier model based on the classification probability distribution of the training sample sets, and then predicts the classification of the test sample sets. On this basis, the proposed model is used to analyze the seasonal distribution difference characteristics of the aerosol types at the SGP station, and the experimental results are consistent with the climatic environment characteristics of this region. In order to verify the accuracy of classification results of the proposed model, the aerosol classification threshold standards are established by using the matching method combining AERONET station data and high spectral resolution lidar profile data. The results show that compared with the traditional aerosol classification algorithm, the aerosol classification results obtained by the proposed model have a high consistency with the results determined based on threshold criteria, which can provide ground data support for aerosol inversion by remote sensing equipment such as satellites.

Key words atmospheric optics; aerosol classification; machine learning; naive Bayesian classifier; aerosol characteristics

1 引言

气溶胶是影响气候变化和地球能量平衡的关键因

素之一^[1-2], 气溶胶的来源可分为自然源和人为源, 气溶胶粒子的大小、形状和成分随来源的不同而不同, 故可将气溶胶划分为多个类型。气溶胶具有高度的时空

收稿日期: 2022-01-20; 修回日期: 2022-03-26; 录用日期: 2022-04-22

基金项目: 国家自然科学基金(61875089, 62175114)

通信作者: *jianhuachang@nuist.edu.cn

可变性,它们通过吸收和散射太阳辐射改变地球的收支平衡从而影响气候,因此研究不同地区、不同季节气溶胶类型的差异对改善气候环境具有重要意义^[3-5]。对气溶胶粒子的研究包括其基本特性、时空分布和传输轨迹等,气溶胶探测手段主要可分为地基观测和卫星遥感探测。太阳光度计作为地基观测的主要设备之一,它能准确探测某一站点的气溶胶物理和光学特性,但由于全球地面观测站点的缺少和观测时间的不连续^[6-7],故基于卫星的遥感观测近年来受到了国内外的广泛关注,它不仅弥补了地基观测站点不足的缺陷,还能更好地连续探测气溶胶高分辨率的时空分布。然而,遥感数据会受到时空分辨率的限制,进而反演得到的气溶胶需要借助地面数据进行验证^[8-10]。

目前在研究气溶胶分类的相关领域中,研究者们多采用聚类分析、阈值设定和机器学习等方法。在聚类分析气溶胶类型方面:Omar等^[11]将AERONET太阳光度计反演的包括气溶胶单次散射反照率、不对称因子、复折射率指数等在内的26个微物理和光学特性作为输入,用K均值聚类获得了6种气溶胶集群;Taylor等^[12]根据全球7年的戈达德化学气溶胶辐射传输(GOCART)光学深度数据用K均值聚类算法得到了10个气溶胶混合物的最优聚类数目,进一步确定了各类气溶胶混合物的地理和季节分布情况。聚类算法对分类结果可解释性较强,但其使用多个带有偏差的气溶胶光学和微物理特性,进而提高了算法的随机误差性。阈值法可以有效地规避这种随机误差,基于少量参数就可实现气溶胶分类:Lee等^[13]根据气溶胶的主要尺寸模式和辐射吸收率,由单次散射反照率和精细模式分数确定了气溶胶的辐射吸收类型;Burton等^[14]研究了由高光谱分辨率激光雷达(HSRL)测得的气溶胶强度参数(激光雷达比、退偏比、后向散射颜色比和光谱去极化比)随气溶胶类型的变化而变化的规律,区分出8种气溶胶类型,附加的光谱去极化比解决了城市和生物质燃烧产生的烟雾气溶胶难以区分的问题。然而,阈值法确定的气溶胶类型在实际应用分析时还需要进一步对气溶胶特性进行复杂的光学和化学测量,后续处理过程繁琐。相较聚类和阈值等传统气溶胶分类方法,机器学习凭借高效稳定和对结果可解释性强等优点近年来被引入到该研究领域。Choi等^[15]开发了一种基于多个卫星数据使用随机森林(RF)模型对气溶胶分类的机器学习新方法,该模型以AERONET数据集为训练依据,可识别7种气溶胶类型,与之前基于卫星数据利用经验阈值分类的结果相比,RF模型具有更高的分类精度。

基于此,本文提出了一种基于朴素贝叶斯分类器(NBC)的气溶胶类型识别模型,该模型选取AERONET站点数据作为训练和测试样本,区分了5种类型气溶胶的季节分布变化。为了检测该模型的分类型精度,以AERONET站点观测时间为基准,筛选对应的HSRL时间廓线数据,并建立基于HSRL数据的气溶胶分类阈值标准来进行验证。与传统算法相比,

所提模型不用考虑气溶胶各输入特征属性之间的关系,通过学习类别条件概率估计构造分类项与类别之间的映射规则,从而做出最优类别判断,因此具有较好的气溶胶分类性能。

2 方法与原理

2.1 AERONET气溶胶光学特性

AERONET是一个全球气溶胶自动检测网络,该网络具有统一数据收集、校准和数据处理的特征。AERONET利用CIMEL光度计同时测量太阳和天空辐射,包括了650多个全球地面分布的监测站点,提供了三个等级的气溶胶观测数据:Level 1(未筛选过的产品)、Level 1.5(云剔除产品)、Level 2(云剔除和质量控制的产品)。

本文提出的NBC气溶胶分类模型使用了5个输入变量,采用2020年6月—2021年5月AERONET的美国俄克拉何马州拉蒙特的南部大平原(SGP)站(36.605°N, 97.486°W)经过质量保证的Level 1.5气溶胶光学反演数据^[16],包括440~870 nm波段的消光Angstrom波长指数(α_{ext})、440~870 nm波段的吸收Angstrom波长指数(α_{abs})、440 nm波长的单次散射反照率(S_{ssa})、440 nm波长的复折射率的实部(m_r)和440 nm波长的复折射率虚部(m_i)。由于Level 2数据量不足,而Level 1.5数据充足且较完整,基于这两点,本文在选取数据时,比对了Level 2数据的所有测量时间点,并以此为标准,适当剔除了Level 1.5数据,使得实验数据在数量充足的同时也有着一定的质量保证。

不同类型的气溶胶有不同的光学和微物理性质,因此了解气溶胶光学和微物理性质的差异有利于区分气溶胶的类型。 α_{ext} 是描述气溶胶消光光学厚度曲线斜率与波长的关系函数^[17],440~870 nm波段的 α_{ext} 的计算公式为

$$\alpha_{\text{ext}} = -\frac{\ln[\lambda_{\text{ext}}(870)] - \ln[\lambda_{\text{ext}}(440)]}{\ln 870 - \ln 440}, \quad (1)$$

式中: $\lambda_{\text{ext}}(\cdot)$ 表示气溶胶消光光学厚度。

α_{abs} 是描述气溶胶吸收光学厚度曲线斜率与波长的关系函数,440~870 nm波段的 α_{abs} 的计算公式为

$$\alpha_{\text{abs}} = -\frac{\ln[\lambda_{\text{abs}}(870)] - \ln[\lambda_{\text{abs}}(440)]}{\ln 870 - \ln 440}, \quad (2)$$

式中: $\lambda_{\text{abs}}(\cdot)$ 表示气溶胶吸收光学厚度。

S_{ssa} 反映了气溶胶的散射系数占总消光系数的比值。 m_r 与 m_i 无关,分别代表了气溶胶的散射和吸收能力, m_r 较高表明气溶胶含水量较低,而 m_i 较高则表明气溶胶吸收率较高。

图1为5个输入变量在SGP站2020年6月—2021年5月期间的季节均值。由图1可知:夏季 α_{ext} 和 α_{abs} 均值最高,这是因为夏季高温高湿的气候会造成粗细模态气溶胶粒子共存;春季 α_{ext} 和 α_{abs} 均值达到最低,表明春季受散射性为主的粗模态气溶胶粒子的影响较大; S_{ssa} 变化较弱,在秋季均值达到最低,夏季达到最高; m_r 与 m_i 的变化趋势与 S_{ssa} 相反,这表明秋季受吸收性为

主的细模态气溶胶粒子影响较大^[18-20]。

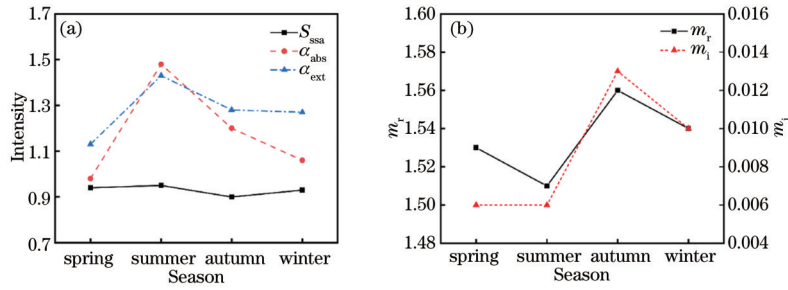


图 1 SGP 站 5 个气溶胶特征参数的季节均值。(a) α_{ext} 、 α_{abs} 和 S_{ssa} 季节均值；(b) m_r 和 m_i 季节均值

Fig. 1 Seasonal mean values of 5 aerosol characteristic parameters at SGP station. (a) Seasonal mean values of α_{ext} , α_{abs} and S_{ssa} ; (b) seasonal mean values of m_r and m_i

2.2 基于 NBC 模型的气溶胶整体识别方法

基于 AERONET 数据利用 NBC 模型实现气溶胶分类的整体流程图如图 2 所示。模型的输入变量分别为 α_{ext} 、 α_{abs} 、 S_{ssa} 、 m_r 和 m_i 。输出的 5 种气溶胶类型分别为城市工业型 (UI)、生物质燃烧型 (BB)、沙尘型 (DU)、海洋型 (MA) 和混合型 (MT)。选取气溶胶输入变量的有效样本给定所属类别标签组成训练集, 计算其类别条件概率生成分类器, 从而输出测试集结果,

其中 NBC 模型识别气溶胶类型是以 Hamill 等^[17]提出的基于任意 AERONET 数据的马氏距离聚类算法得到的 5 类气溶胶类型参考簇质心为训练依据的, 如表 1 所示。最后, 根据大气辐射测量 (ARM) 计划提供的由 HSRL 观测的气溶胶退偏比 (δ_a)、激光雷达比 (S_a) 和 532 nm/1064 nm 后向散射颜色比 (χ_a), 建立阈值标准划分各季节气溶胶类型以验证上述基于 NBC 模型的输出结果的正确性。

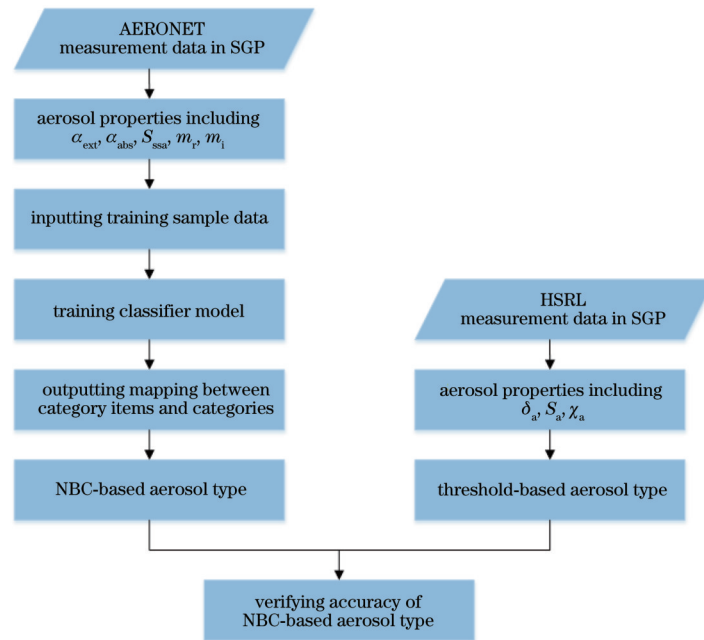


图 2 整体算法流程图

Fig. 2 Flow chart of overall algorithm

表 1 5 种气溶胶参考簇质心

Table 1 Centroids of five aerosol reference clusters

Aerosol type	α_{ext}	α_{abs}	S_{ssa}	m_r	m_i
Urban industry (UI)	1.76	1.15	0.96	1.40	0.005
Biomass burning (BB)	1.87	1.30	0.89	1.48	0.020
Dust (DU)	0.28	1.75	0.91	1.47	0.004
Marine (MA)	0.59	0.93	0.97	1.40	0.001
Mixed (MT)	1.32	1.20	0.92	1.45	0.011

2.3 基于 NBC 的气溶胶分类模型

在气溶胶分类研究中, NBC 模型是根据统计概率知识生成分类器的, 得到的气溶胶类型可解释性强。基于地基或卫星遥感等观测设备探测的气溶胶数据常存在缺失或者异常等问题, 该模型对缺失数据不敏感的特点可有效减小分类结果的误差。此外, 大气中的气溶胶成分复杂、类别多样且不断变化, 具有计算效率高、实时性强和在处理多类别问题表现优异的特点的

NBC 模型更适用于这一分类场景。此外, NBC 模型在有监督的环境下高效学习, 它不需要使用大量参数, 进而对于小训练集的气溶胶样本, 其也能获得较好的分类结果^[21-22]。

NBC 模型的工作原理是应用具有独立性假设的贝叶斯定理, 即在给定类别 C 的条件下, 假设各特征属

性之间是相互独立的^[23]。NBC 模型的具体流程如图 3 所示。首先, 分别给定一组训练集和待分类数据集。然后, 根据训练集计算类别条件概率和类别先验概率, 再通过贝叶斯公式来计算类别后验概率。最后, 取后验概率最大时的类作为待分类数据项所属的类。

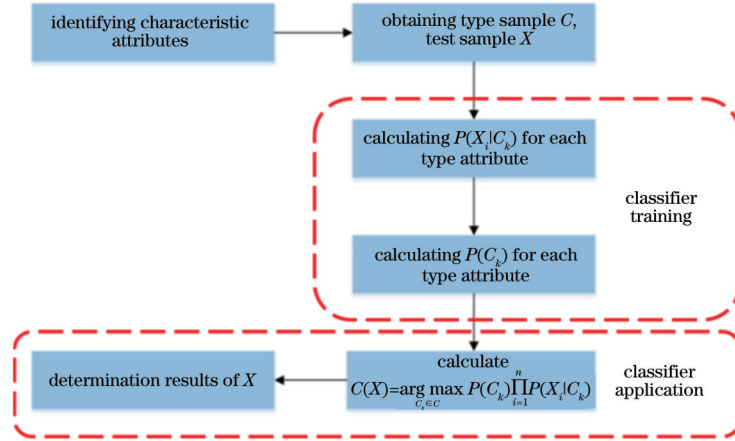


图 3 NBC 模型流程图

Fig. 3 Flow chart of NBC model

通过构造 NBC 分类模型预测 2020 年 6 月 - 2021 年 5 月期间 SGP 站的 5 种气溶胶类型, 获取 5 个有效输入变量 α_{ext} 、 α_{abs} 、 S_{ssa} 、 m_r 和 m_i , 共 845 组样本, 分别选取 25%、50%、75% 的数据作为训练集样本得到 5 类气溶胶的总体识别率为 90.7%、91.1%、91.9%。为了突显 NBC 模型对小训练集样本也有着较好的分类性能这一优势和提高训练效率(在不同大小的训练集样本中总体识别精度相差不大), 本文选取 25% 的数据组成训练集用于训练, 其余样本组成测试集, 具体分类步骤如下。

1) 设气溶胶输入属性随机变量集为 $X = \{X_1, X_2, X_3, X_4, X_5\}$, 类别变量集为 $C = \{c_1, c_2, c_3, c_4, c_5\}$, 其中 c_1 为城市工业型气溶胶、 c_2 为生物质燃烧型气溶胶、 c_3 为沙尘型气溶胶、 c_4 为海洋型气溶胶、 c_5 为混合型气溶胶取值范围。 X_i 表示用于 NBC 模型输入的 5 种气溶胶特征参数, x_i 表示每种气溶胶特征参数的个数, 故 x_i 为属性 X_i 的取值, 可设测试集中气溶胶样本为 $x_i = \{x_1, x_2, \dots, x_{633}\}$ 。

2) 由贝叶斯公式得到样本 x_i 属于 c_j ($j = 1, 2, \dots, 5$) 的概率为

$$P(C = c_j | x_1, x_2, \dots, x_{633}) = \frac{P(c_j)P(x_1, x_2, \dots, x_{633} | c_j)}{P(x_1, x_2, \dots, x_{633})} = \frac{aP(c_j)P(x_1, x_2, \dots, x_{633} | c_j)}{aP(c_j)P(x_1, x_2, \dots, x_{633} | c_j)}, \quad (3)$$

式中: a 为正则化因子。先验概率 $P(c_j)$ 和条件概率 $P(x_1, x_2, \dots, x_{633} | c_j)$ 的值都可通过训练集样本估算, 后验概率 $P(c_j | x_1, x_2, \dots, x_{633})$ 最大时的 c 便是给定的属性 x_i 对应的类别, 后验概率由链式法可得, 即

$$P(c_j | x_1, x_2, \dots, x_{633}) = aP(c_j) \prod_{i=1}^{633} P(x_i | x_1, \dots, x_{i-1}, c_j). \quad (4)$$

对于给定的训练集样本总数 212 组, 均有先验概率 $P(c_j) = s_j / 212$, s_j 是类别 c_j 在训练集样本的实例数。条件概率 $P(x_i | c_j)$ 的计算过程如下。

1) 当 X_i 为离散变量时, 条件概率为 $P(x_i | c_j) = s_{ji} / 212$, 其中 s_{ji} 是在属性 X_i 上具有值 x_i 的类 c_j 的训练样本实例数。

2) 当 X_i 为连续变量时, 条件概率为

$$P(x_i | c_j) = g(x_i, \mu_{c_j}, \sigma_{c_j}) = \frac{1}{\sqrt{2\pi\sigma_{c_j}^2}} \exp\left[-\frac{(x_i - \mu_{c_j})^2}{2\sigma_{c_j}^2}\right], \quad (5)$$

式中: $g(x_i, \mu_{c_j}, \sigma_{c_j})$ 是属性 X_i 的高斯密度函数; μ_{c_j} 和 σ_{c_j} 分别为平均值和标准方差。

利用生成的 NBC 模型实现对测试集气溶胶样本的识别, 通过对比模型识别结果与给定分类结果得到, 633 组气溶胶样本有 574 组识别正确, NBC 模型总体识别率为 91%, 5 类气溶胶的个数 (N) 分别为 119、106、155、66 和 187。59 组样本被误判的具体情况如表 2 所示, 城市工业型、生物质燃烧型、沙尘型、海洋型和混合型气溶胶的识别率分别为 88%、96%、92%、95% 和 87%。

表 3 为 NBC 模型确定的各类气溶胶特性的平均值。从表 3 可以看出, 该模型能较好地区分 5 种气溶胶不同的光学特性, 其中: 城市工业型与海洋型气溶胶的 S_{ssa} 、 m_r 和 m_i 的值的分布相似; 混合型与生物质燃烧型气溶胶的 α_{abs} 的值的分布相似, 但该值的分布与沙尘型气溶胶的 S_{ssa} 和 m_r 的值的分布相似, 故容易造成错误

识别,导致城市工业型气溶胶和混合型气溶胶的误判率较高。造成这两类气溶胶误判率较高的原因可以认为是气溶胶被错归为具有类似微物理性质和光学性质

的类,这在 Omar 等^[11]利用 K 均值聚类算法和 Choi 等^[15]提出的使用随机森林模型进行气溶胶分类的实验结果中都有所表现。

表 2 各类气溶胶误判情况
Table 2 Misjudgment of various aerosols

Aerosol type	UI	BB	DU	MA	MT	Actual number	Recognition rate /%
UI	105	0	6	7	1	119	88
BB	0	102	0	0	4	106	96
DU	9	0	142	0	4	155	92
MA	3	0	0	63	0	66	95
MT	0	17	8	0	162	187	87
Test number	117	119	156	70	171	633	

表 3 基于 NBC 模型确定的各类气溶胶特性的平均值
Table 3 Average values of various aerosol characteristics determined by NBC model

Aerosol characteristic	UI	BB	DU	MA	MT
N	119	106	155	66	187
α_{ext}	1.340	1.380	0.840	0.960	1.210
α_{abs}	1.050	1.190	1.370	0.990	1.160
S_{ssa}	0.960	0.860	0.930	0.980	0.920
m_r	1.510	1.560	1.550	1.510	1.530
m_i	0.003	0.022	0.006	0.001	0.011

3 实验结果与分析

3.1 不同类型气溶胶的季节特征

为了分析气溶胶类型随季节的变化特征,利用训练好的 NBC 模型得出了 SGP 站四季气溶胶类型的占比分布,图 4 为不同季节气溶胶类型的占比分布图。从图 4 中可以看出,全年混合型气溶胶占比最高,5 类气溶胶的占比差异随季节的变化较明显。SGP 深处内陆降水少,多风沙天气。春季受东南风吹送,沙尘现象出现次数较多,沙尘型气溶胶在春季占主导地位,占比为 49%,同时春季农业机械生产活动多且农业机械化程度高,进而城市工业性气溶胶也占有一定比例,占比为 31%。SGP 的亚热带大陆性气候导致全年降水量少且集中在夏季,因此夏季大气中的吸湿性气溶胶占比^[24-25]增加,其中混合型气溶胶占比最高(32%),海洋性气溶胶相比其他季节占比也最高,占比为 23%。秋季农业生产增多,生物质燃烧等活动频繁,导致生物质燃烧型气溶胶占据主导地位,占比为 35%,混合型气溶胶占比略低(34%)。冬季受山风、海风和极地强冷空气的影响,常伴有大雪和由风暴带来的沙尘,导致沙尘气溶胶含量多,占比最高为 29%,同时冬季各类气溶胶占比相对均衡。各季气溶胶分布结果符合 SGP 地区实际气候环境特征。

3.2 HSRL 数据验证

本文采用 ARM 计划中 HSRL^[26]在 SGP 站同时段

探测到的气溶胶光学特性数据来验证上述气溶胶类型分布结果,基于 δ_a 、 S_a 和 χ_a 建立阈值标准对气溶胶分类,其中: δ_a 反映了气溶胶粒子的非球形程度^[27],球形粒子的 δ_a 值最小; S_a 是气溶胶消光系数与后向散射系数的比值,吸收性越强的气溶胶粒子, S_a 越大; χ_a 通常与气溶胶粒子大小成反比。反演的气溶胶光学参数有效高度一般从 900 m 处开始,故在处理数据时均选择 900 m 作为起始高度。本文在参考已有的利用 HSRL 的 δ_a 和 S_a 设定阈值标准^[14]分类不同类型气溶胶的基础上,考虑到该阈值标准对城市工业型和生物质燃烧型气溶胶的划分有较多重叠部分,故结合 χ_a ^[28]对这两种气溶胶进行区分,建立的 5 类气溶胶阈值标准:

- 1) 当 $5\% < \delta_a < 10\%$, $36 \text{ sr} < S_a < 75 \text{ sr}$, $\chi_a < 3$ 时,气溶胶类型为城市工业型;
- 2) 当 $2\% < \delta_a < 9\%$, $30 \text{ sr} < S_a < 86 \text{ sr}$, $\chi_a > 3$ 时,气溶胶类型为生物质燃烧型;
- 3) 当 $17\% < \delta_a < 35\%$, $28 \text{ sr} < S_a < 58 \text{ sr}$ 时,气溶胶类型为沙尘型;
- 4) 当 $1\% < \delta_a < 10\%$, $13 \text{ sr} < S_a < 25 \text{ sr}$ 时,气溶胶类型为海洋型;
- 5) 其余气溶胶为混合型气溶胶。

以 HSRL 在 2021 年 3 月 29 日 14:00—19:00 期间观测到的气溶胶为例,图 5 为在该时间段内的气溶胶退偏比时空分布情况。可以看出,在垂直高度为 0.9~2.0 km,时间为 16:00—19:00 范围内 δ_a 集中在 17%~22% 之间,此时可初步判定该层为沙尘气溶胶层。然后,以该观测时间段内两个不同的时刻 15:00(未出现沙尘气溶胶)和 18:00(出现沙尘气溶胶)为例,分析利用信号反演得到的 δ_a 、 S_a 和 χ_a 的分时廓线结果,如图 6 所示。对比两个时刻各气溶胶光学特性参数廓线可以看出,在高度为 0.9~2.0 km 范围内,相比 15:00 的气溶胶光学特性值,18:00 出现的气溶胶具有相对较高的 δ_a 值(基本在 0.17~0.22 之间),有较低的 S_a 值(基本在 35~40 sr 之间),有较低的 χ_a 值(基本在 1.3~1.7 之间),总体与沙尘气溶胶特征吻合。此外,从图 6(a) 的 18:00 的廓线中可以看出: δ_a 最大值达到 0.22,属于初等程度的沙尘气溶胶,说明该沙尘气溶胶层中非球

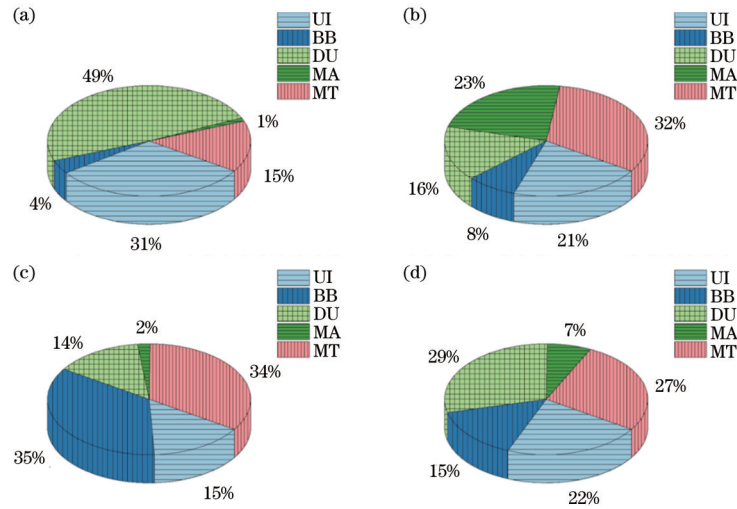


图 4 NBC 模型对 2020 年—2021 年 SGP 站四季的气溶胶分类结果占比。(a) 春季分类占比; (b) 夏季分类占比; (c) 秋季分类占比; (d) 冬季分类占比

Fig. 4 Proportion of aerosol classification results of SGP station in four seasons from 2020 to 2021 obtained by NBC model. (a) Proportion of spring classification; (b) proportion of summer classification; (c) proportion of autumn classification; (d) proportion of winter classification

形粒子所占比例偏少;在沙尘气溶胶层顶部(1.7~2.0 km)处, δ_a 值随高度的增加而减小,这与层内气溶胶的垂直分布结构差异有关,原因可能是重力作用使得较小尺度的沙尘粒子停留在了层顶部中。同样地,

在未出现沙尘气溶胶前,结合图 5 和图 6 分析可得,在高度为 0.9~3.0 km,时间为 14:00—16:00 范围内存在着混合型气溶胶与城市工业型气溶胶。

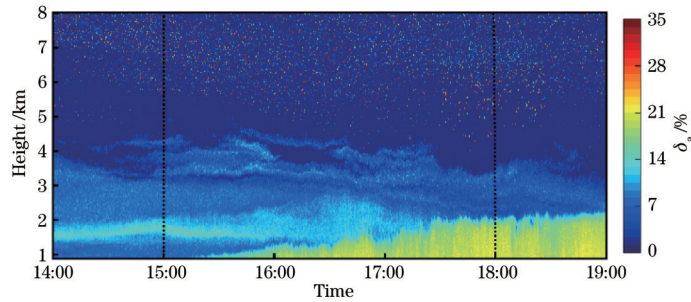


图 5 HSRL 沙尘型气溶胶退偏比分布(2021 年 3 月 29 日 14:00—19:00)

Fig. 5 Depolarization ratio distribution of HSRL dust aerosol (from 14:00 to 19:00 on March 29, 2021)

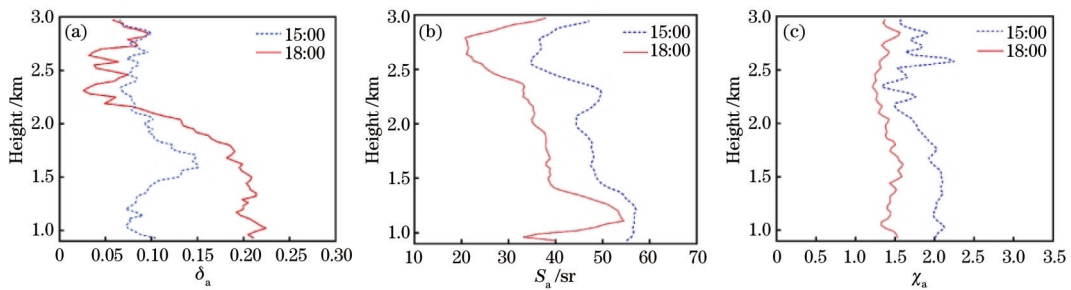


图 6 HSRL 气溶胶光学参数分时廓线结果。(a) 退偏比廓线; (b) 激光雷达比廓线; (c) 色比廓线

Fig. 6 Time-sharing profile results of HSRL aerosol optical parameters. (a) Depolarization ratio profile; (b) lidar ratio profile; (c) color ratio profile

为了清晰直观地看出 HSRL 观测到的气溶胶不同季节的垂直结构分布,选取了 2020 年 8 月—2021 年 5 月 4 个季节各典型月份 δ_a 的观测结果,分别如图 7 所

示。其中,由于 HSRL 在夏季实际观测天数的缺少,故选择 2020 年 8 月 25 日—2020 年 9 月 24 日这一时间段内的 δ_a 分布。可以看出,图 7(a) 中 2021 年 3 月 1 日—

31日和图7(d)中2021年1月1日-31日的 δ_a 值偏大,主要集中在17%~33%之间,表明沙尘等非球形气溶胶粒子多。图7(b)中2020年8月25日-2020年9月

24日和图7(c)中2020年11月1日-30日的 δ_a 值明显降低,主要集中在5%~15%之间,球形气溶胶粒子多。

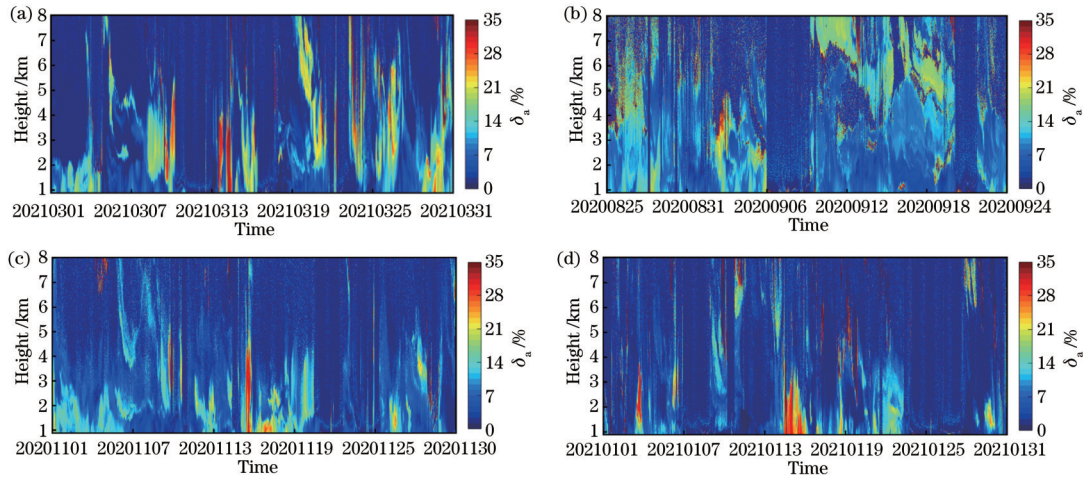


图7 2020年-2021年HSRL在SGP站4个季节典型月份的气溶胶退偏比观测结果。(a)2021年3月1日-31日;(b)2020年8月25日-9月24日;(c)2020年11月1日-30日;(d)2021年1月1日-31日

Fig. 7 Observation results of aerosol depolarization ratio of HSRL in typical months of four seasons at SGP station from 2020 to 2021. (a) March 1 to 31, 2021; (b) August 25 to September 24, 2020; (c) November 1 to 31, 2020; (d) January 1 to 31, 2021

图8为HSRL观测到的气溶胶 δ_a 、 S_a 和 χ_a 的季节平均垂直分布廓线。从图8(a)可以看出, δ_a 的季节均值是 $\delta_{a, spring} > \delta_{a, winter} > \delta_{a, autumn} > \delta_{a, summer}$ 。从图8(b)可以看出, S_a 的季节均值是 $S_{a, autumn} > S_{a, spring} > S_{a, summer} >$

$S_{a, winter}$ 。从图8(c)可以看出, χ_a 的季节均值是 $\chi_{a, winter} > \chi_{a, autumn} > \chi_{a, summer} > \chi_{a, spring}$ 。由各类气溶胶的阈值范围可知,由HSRL确定的各季节主要类型气溶胶与NBC模型输出结果一致。

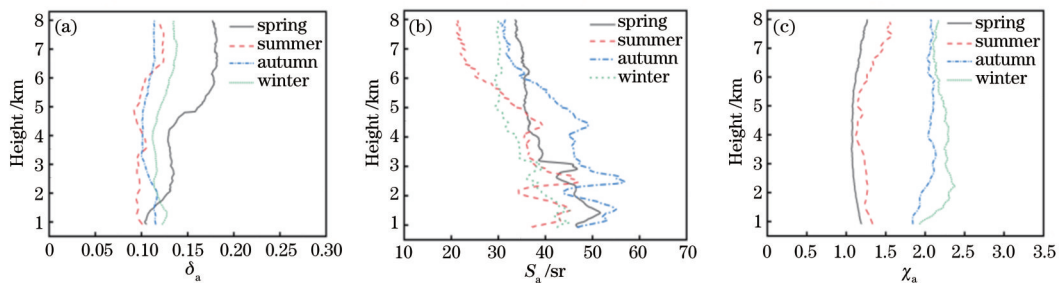


图8 2020年-2021年高分辨率激光雷达SGP站气溶胶光学参数季节平均。(a)退偏比季节平均;(b)激光雷达比季节平均;(c)色比季节平均

Fig. 8 Seasonal average of aerosol optical parameters of high resolution lidar SGP station from 2020 to 2021. (a) Seasonal average of depolarization ratio; (b) seasonal average of lidar ratio; (c) seasonal average of color ratio

为了进一步分析分类结果的准确性,取HSRL观测高度为0.9~8.0 km,以AERONET地面站点观测时间为基准,在它的-15~15 s时间范围内,将HSRL与相应AERONET的一次观测时间最接近的情况认为是一个匹配点,可近似认为两者观测到的为同一气溶胶团。针对2020年5月-2021年6月SGP站的HSRL观测数据共筛选出581根廓线,即581个匹配点。

以上述气溶胶阈值分类标准对581个匹配点进行气溶胶分类的结果为参考,来验证NBC模型的识别准确性。此外,还将基于AERONET数据的马氏距离聚类^[17]气溶胶分类算法作为对比,三者的分类结果如表

4所示。表4显示,相比马氏距离聚类算法结果,所提NBC模型与基于阈值标准划分的气溶胶类型总体表现出更相近的分布占比,具有更准确的分类结果。从表4中还可以看出,马氏距离聚类算法对混合型气溶胶的检测比例较高,各季节的主要气溶胶类型均为混合型气溶胶,气溶胶类型随季节性变化特点不明显。

为了检测少数缺失和异常数据对实验结果的影响,利用NBC模型和马氏距离聚类算法对2020年6月-2021年5月内31个缺失和异常值所在的31组气溶胶光学特性样本进行分类实验。实验结果表明,在对应时间点上,由NBC模型得到的24组气溶胶类型与基于HSRL阈值标准确定的结果相一致,但马氏距

离聚类算法只有 16 组与之一致,可见类别的数据属性失衡会导致聚类算法效果变差,而 NBC 模型对缺失数据不敏感的特点在气溶胶类型的识别研究上更有优势。然而,由表 4 可知,NBC 模型对混合型和城市工业

型气溶胶的识别与基于 HSRL 阈值标准划分的结果占比差异较大,根据上述对各类气溶胶光学特性平均值的分析,不难发现这是普遍存在的具有相近光学和微物理性质类的气溶胶混淆识别造成的。

表 4 不同分类算法得出的气溶胶类型四季占比情况

Table 4 Proportion of aerosol types in four seasons obtained by different classification algorithms unit: %

Classification algorithm	Aerosol type	Spring	Summer	Autumn	Winter
NBC	Urban industry	31	21	15	22
	Biomass burning	4	8	35	15
	Dust	49	16	14	29
	Marine	1	23	2	10
	Mixed	15	32	34	24
Mahalanobis distance	Urban industry	23	18	12	21
	Biomass burning	2	6	23	10
	Dust	29	9	11	20
	Marine	1	17	2	6
	Mixed	45	50	52	43
Threshold criteria	Urban industry	27	14	10	17
	Biomass burning	5	6	31	11
	Dust	43	13	15	33
	Marine	4	28	4	9
	Mixed	21	39	40	30

4 结 论

将机器学习应用到 AERONET 数据的气溶胶分类研究中,提出了一种基于 NBC 的气溶胶类型识别模型,通过计算气溶胶各类别的概率分布生成分类器模型,输出了 5 种类型气溶胶(城市工业型、生物质燃烧型、沙尘型、海洋型和混合型)。实验结果表明:该地区全年混合型气溶胶最多;春冬季以沙尘型气溶胶为主,受沙尘影响,粗粒子含量高;夏季中海洋型气溶胶相比其他季节占比更大,受粗细粒子共同影响;秋季以生物质燃烧型气溶胶为主,细粒子含量高。通过 HSRL 测得的 δ_a 、 S_a 和 χ_a 建立新的阈值标准判定了 5 种类型气溶胶,对所提气溶胶分类模型的识别结果进行了验证。与传统气溶胶分类算法相比,NBC 模型的分类结果与基于 HSRL 阈值标准确定的结果更具一致性,能更明显准确地体现气溶胶类型的季节变化特征,可为卫星等遥感设备反演的气溶胶产品提供地面数据支撑。然而,所提模型对城市污染型和混合型气溶胶的检测精度相对较低,主要原因是气溶胶容易被错归为具有类似微物理性质和光学性质的类,后续工作将深入研究不同类型气溶胶的微物理和光学性质。

参 考 文 献

- [1] Noh Y, Müller D, Lee K, et al. Depolarization ratios retrieved by AERONET sun-sky radiometer data and comparison to depolarization ratios measured with lidar [J]. *Atmospheric Chemistry and Physics*, 2017, 17(10): 6271-6290.
- [2] She L, Xue Y, Guang J, et al. Aerosol optical and physical properties over Beijing[C]//IEEE International Geoscience and Remote Sensing Symposium, July 23-28, 2017, Fort Worth, TX, USA. New York: IEEE Press, 2017: 5962-5965.
- [3] Kumar K R, Sivakumar V, Reddy R R, et al. Identification and classification of different aerosol types over a subtropical rural site in Mpumalanga, South Africa: seasonal variations as retrieved from the AERONET sunphotometer[J]. *Aerosol and Air Quality Research*, 2014, 14(1): 108-123.
- [4] Tan C H, Zhao T L, Xu X F, et al. Climatic analysis of satellite aerosol data on variations of submicron aerosols over East China[J]. *Atmospheric Environment*, 2015, 123: 392-398.
- [5] 麻晓敏,陶宗明,单会会,等.激光雷达探测对流层气溶胶消光系数的统计分布[J]. *光学学报*, 2020, 40(11): 1101003.
Ma X M, Tao Z M, Shan H H, et al. Statistical distribution of extinction coefficients of tropospheric aerosols detected by lidar[J]. *Acta Optica Sinica*, 2020, 40(11): 1101003.
- [6] Torres O, Ahn C, Chen Z. Improvements to the OMI near-UV aerosol algorithm using A-train CALIOP and AIRS observations[J]. *Atmospheric Measurement Techniques*, 2013, 6(11): 3257-3270.
- [7] Lin J Y, Zheng Y, Shen X Y, et al. Global aerosol classification based on Aerosol Robotic Network

- (AERONET) and satellite observation[J]. Remote Sensing, 2021, 13(6): 1114.
- [8] Li X K, Zhang C R, Li W D, et al. Exploring the trend, prediction and driving forces of aerosols using satellite and ground data, and implications for climate change mitigation[J]. Journal of Cleaner Production, 2019, 223: 238-251.
- [9] Filonchyk M, Yan H, Zhang Z, et al. Combined use of satellite and surface observations to study aerosol optical depth in different regions of China[J]. Scientific Reports, 2019, 9: 6174.
- [10] Bibi H, Alam K, Chishtie F, et al. Intercomparison of MODIS, MISR, OMI, and CALIPSO aerosol optical depth retrievals for four locations on the Indo-Gangetic plains and validation against AERONET data[J]. Atmospheric Environment, 2015, 111: 113-126.
- [11] Omar A H, Won J G, Winker D M, et al. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements[J]. Journal of Geophysical Research, 2005, 110(D10): D10S14.
- [12] Taylor M, Kazadzis S, Amiridis V, et al. Global aerosol mixtures and their multiyear and seasonal characteristics [J]. Atmospheric Environment, 2015, 116: 112-129.
- [13] Lee J, Kim J, Song C H, et al. Characteristics of aerosol types from AERONET sunphotometer measurements[J]. Atmospheric Environment, 2010, 44(26): 3110-3117.
- [14] Burton S P, Ferrare R A, Hostetler C A, et al. Aerosol classification using airborne High Spectral Resolution Lidar measurements: methodology and examples[J]. Atmospheric Measurement Techniques, 2012, 5(1): 73-98.
- [15] Choi W, Kang H, Shin D, et al. Satellite-based aerosol classification for capital cities in Asia using a random forest model[J]. Remote Sensing, 2021, 13(13): 2464.
- [16] 梁天辰, 孙林, 王永吉. 基于深度学习反演区域气溶胶光学厚度[J]. 光学学报, 2021, 41(4): 0401002.
Liang T C, Sun L, Wang Y J. Retrieval of regional aerosol optical depth using deep learning[J]. Acta Optica Sinica, 2021, 41(4): 0401002.
- [17] Hamill P, Giordano M, Ward C, et al. An AERONET-based aerosol classification using the Mahalanobis distance[J]. Atmospheric Environment, 2016, 140: 213-233.
- [18] Li J, Carlson B E, Laciš A A. Using single-scattering albedo spectral curvature to characterize East Asian aerosol mixtures[J]. Journal of Geophysical Research, 2015, 120(5): 2037-2052.
- [19] Chung C E, Kim S W, Lee M, et al. Carbonaceous aerosol AAE inferred from *in situ* aerosol measurements at the Gosan ABC super site, and the implications for brown carbon aerosol[J]. Atmospheric Chemistry and Physics, 2012, 12(14): 6173-6184.
- [20] Müller T, Schladitz A, Kandler K, et al. Spectral particle absorption coefficients, single scattering albedos and imaginary parts of refractive indices from ground based *in situ* measurements at Cape Verde Island during SAMUM-2[J]. Tellus B, 2011, 63(4): 573-588.
- [21] Lü T, Yan P, Yuan H W, et al. Spam filter based on naive Bayesian classifier[J]. Journal of Physics: Conference Series, 2020, 1575(1): 012054.
- [22] Feng C X, Zhao H L, Li Y E, et al. Improved detection of focal cortical dysplasia in normal-appearing FLAIR images using a Bayesian classifier[J]. Medical Physics, 2021, 48(2): 912-925.
- [23] Sun L J, Lu Y, Jin J G, et al. An integrated Bayesian approach for passenger flow assignment in metro networks[J]. Transportation Research Part C: Emerging Technologies, 2015, 52: 116-131.
- [24] 赵佳佳, 顾芳, 张加宏, 等. 单颗粒气溶胶的吸湿增长模型及散射特性研究 [J]. 光学学报, 2020, 40(5): 0301001.
Zhao J J, Gu F, Zhang J H, et al. Hygroscopic growth model and scattering characteristics of single-particle aerosols[J]. Acta Optica Sinica, 2020, 40(5): 0301001.
- [25] Zhang L, Sun J Y, Shen X J, et al. Observations of relative humidity effects on aerosol light scattering in the Yangtze River Delta of China[J]. Atmospheric Chemistry and Physics, 2015, 15(14): 8439-8454.
- [26] Cheng Z T, Liu D, Luo J, et al. Field-widened Michelson interferometer for spectral discrimination in high-spectral-resolution lidar: theoretical framework[J]. Optics Express, 2015, 23(9): 12117-12134.
- [27] 张小林. 沙尘气溶胶粒子模型的线退偏比特性[J]. 光学学报, 2016, 36(8): 0829001.
Zhang X L. Linear depolarization ratios characteristics of dust aerosol particles model[J]. Acta Optica Sinica, 2016, 36(8): 0829001.
- [28] Groß S, Esselborn M, Weinzierl B, et al. Aerosol classification by airborne high spectral resolution lidar observations[J]. Atmospheric Chemistry and Physics, 2013, 13(5): 2487-2505.