

黄渤海晴空像元下雾霾识别的机器学习算法

沈晓晶, 姜祝辉, 陈 建, 刘 娟, 马 明
(北京应用气象研究所 北京 100029)

摘要: 为实时有效地识别和监测雾霾, 选取 2013 年 10 月 4 日与 10 月 5 日的两景 MODIS 遥感图像, 利用不同类型像元的光谱差异, 基于神经网络模型, 建立海上雾霾的遥感识别机器学习算法。经检验表明, 在晴空条件下, 机器学习算法反演精度相对较高, 已超过 98%, 利用算法可以实现对中国近海上空雾霾的监测。

关键词: 雾霾; 机器学习; 晴空

中图分类号: P733.3

文献标识码: A

文章编号: CN11-1780(2021)03-0067-05

Machine learning algorithm of haze detection over the Yellow and Bohai Sea

SHEN Xiaojing, JIANG Zhuhui, CHEN Jian, LIU Juan, MA Ming

(Institute of Applied Meteorology in Beijing, Beijing 100029, China)

Abstract: In order to identify and monitor the haze effectively, in this paper, we selected two MODIS images on 4th and 5th October, 2013. Based on spectral differences between different kinds of images and neural network model, a machine learning algorithm for identification of marine haze was established in this paper. By validation, the accuracy of the algorithm was more than 98%. With this haze detection algorithm, we can monitor haze for long time over East China Seas.

Key words: Haze; Machine learning; Sunny

引 言

生态文明建设概念的提出, 给人与自然和谐相处这一话题带来了越来越多的关注。雾霾是中国近海地区最为典型的气溶胶污染天气类型之一。近几年来, 雾霾天气出现的频率越来越高, 在出行安全、身体健康以及生态环境等方面对人类造成了极大的影响。对气溶胶的研究已成为全球及区域大气环境研究的一个热点。

雾霾是黄渤海区域典型的气溶胶现象, 雾霾天气通常与低的空气能见度相伴发生, 这给机场、轮渡等交通运输造成了很大的安全隐患。雾霾属于灾害性天气, 它会引起环境效应和气候效应等问题, 受到政府部门、科学家以及社会的广泛关注, 成为热点话题。因此, 对雾霾的监测非常必要。

海洋作为全球气候系统的重要组成部分, 对全球气候的变化有重要影响。我国濒临太平洋, 是海洋大国, 拥有 18 000 km 的大陆岸线和 14 000 km 的岛屿岸线。同时, 东部沿海地区是我国经济社会发展相对较快的区域, 也是我国主要的经济贡献区域, 全国约 60%以上的人口都分布在东部沿海^[1]。因此, 海洋对我国的经济社会和国防建设发展有极其重要的战略意义。

传统的地面观测手段费时费力, 且观测的空间覆盖范围有限。自 20 世纪 90 年代开始, 相继有学者利用卫星遥感的手段监测雾霾。HUSAR^[2]基于 AVHRR 的气溶胶光学厚度产品的分布特征, 认为卫星产品在大气气溶胶以及气候相关的研究中起到重要作用。SIFAKIS 等^[3]开发 SIPHA (Satellite Image Processing for Haze and Aerosol Mapping) 算法制作霾和气溶胶的卫星图像, 但该算法存在一些缺点, 需要无云的晴空影像作为参考影像。兰措等^[4]基于 NOAA 卫星云图, 利用不同类型像元的反照率、亮温差对西宁市上空的雾霾进行识别监测。ZHANG 等^[5]基于 TM/ETM 的第一和第三 (红蓝) 两个不同波段的不同类型像元光谱特征数据, 分别针对晴空像元和有雾霾存在的两种天气情况, 提出了分别提取雾霾和薄云信息的监测方法, 并实现了剔除遥感影像中的雾霾信息, 但是该方法在有水体、雪和裸露地表等

地物类型存在时不适用。随着卫星遥感的飞速发展, 卫星遥感在大气气溶胶分布特性以及大气颗粒物浓度的估算等方面的应用越来越多, 但前人的研究主要集中在陆地, 对海上雾霾的监测识别相对很少。本文的重点在于用遥感和机器学习相结合的手段, 建立一种智能识别与监测海上雾霾的算法, 避免经验和半经验算法的数学模型造成的算法误差, 从而更好地服务于海上运输、海气相互作用、气候变化研究等工作。

1 研究区域与数据

1.1 研究区域

黄渤海位处中国近海, 受陆地影响较大, 该区域的海气相互作用、海陆相互作用较明显, 因此, 研究黄渤海区域的雾霾对未来研究海陆相互作用的工作有重要的意义。本文将选取黄渤海作为典型研究区域, 对该区域上空的雾霾做遥感识别算法研究。如图 1 所示, 黄渤海属于半封闭海区, 平均水深大约 31 m, 覆盖面积约 457 000 平方公里^[6]。其跨越的经纬度范围大概在 31°N 至 41°N, 117°E 至 126°E 之间, 通过东海与太平洋相连^[7]。由于近岸与离岸浊度的变化, 该区域水体类型与水质有较大的变化, 水体光谱的变化也较为明显^[8]。

1.2 数据

MODIS 传感器是美国国家航空航天局 (NASA) 地球观测系统 (EOS) 系列卫星的重要传感器之一, 是新一代“图谱合一”的光学传感器^[9]。由于 MODIS 具有较高的时空分辨率, 且免费对外开放, 因此, 在全球范围内得到了广泛应用^[10]。结合黄渤海附近陆地上已有的雾霾天气预报资料以及 NOAA 提供的 MODIS 真彩色遥感图像, 以此为据、通过目视解译的手段来选取雾霾像元, 最终选取的晴空像元为 2013 年 10 月 4 日和 2013 年 10 月 5 日的两景 MODIS-Aqua 图像, 如图 2 所示。本文使用的卫星遥感数据来自海洋水色数据网站 OceanColor Web (<https://oceancolor.gsfc.nasa.gov/>) 的 MODIS L1B 数据。

对所选遥感图像经过去瑞利校正处理后得到去瑞利遥感反射率 R_{rc} , 通过目视解译的方法、ENVI 提取感兴趣区域的手段, 最终获得清洁水体的样本数为 9 217, 浑浊水体的样本数为 8 194, 雾霾的样本数为 9 592。将样本数据分为两部分, 第一部分取约 30% 作为模型训练样本, 剩下约 70% 的部分作为验证样本。

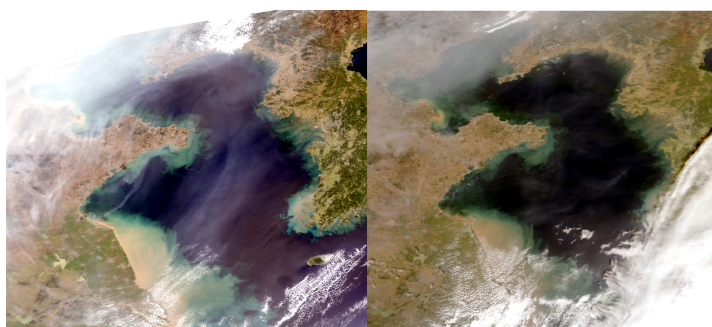


图 1 雾霾遥感识别区域

Fig. 1 Haze detection area

图 2 选取的两景 MODIS 晴空像元真彩色图像

Fig. 2 Two selected images used for haze detection from MODIS

2 方法

2.1 光谱特征参数提取

文中雾霾遥感识别算法建立的基础是利用不同类型像元的光谱特征差异, 基于神经网络模型建立一种机器学习算法。

首先, 需要利用 SeaDAS (SeaWiFS Data Analysis System) 软件对 MODIS L1B 数据进行处理, 得到去瑞利校正后的遥感反射率 R_{rc} , 用于雾霾识别的机器学习算法建立。由于遥感图像上云掩膜的情况比较严重, MODIS 的遥感反射率 R_{rs} 有很多缺省值, 光谱特征差异明显, 因此, 本文选取 R_{rc} 作为光谱特征参数。SeaDAS 是一个开源系统, 是专业处理海洋水色卫星数据的图像综合分析软件包, 可以对各个级别的 MODIS、SeaWiFS 等卫星数据产品 (L0、L1A、L1B、L2、L3Binned、L3SMI) 以及辅助的气象

和臭氧数据做显示、处理、分析和质量控制等处理。本文中使用的 SeaDAS7.5 版本。

其次,利用 ENVI 软件提取遥感图像中的感兴趣区域光谱,提取区域光谱时共分三种类型:清洁水体像元、浑浊水体像元和雾霾像元。具体步骤是先通过目视解译的方法判别出不同类型的像元,包括前文提到的三种类型像元。如图 2 中所示,渤海的浅灰色像元通过目视解译为雾霾,靠近苏北浅滩的黄棕色像元通过目视解译为浑浊水体,渤海海峡以及黄海大部分区域的深绿色像元通过目视解译为清洁水体。再利用 ENVI 软件分别对三种类型的像元进行样本提取,即提取三种不同类型像元的可见光波段光谱。在样本提取时,选择的均是特征较明显的像元,尽量避开边界情况,以保证所选取的样本能够代表该类型的像元。最终提取了 9 217 条清洁水体样本,8 194 条浑浊水体样本,9 592 条雾霾样本。然后,将选取出的样本分为两部分,比例分别为 70%和 30%。其中,30%用于模型训练,70%用于模型验证。

2.2 神经网络模型介绍

随着人工智能 AI (Artificial Intelligence) 技术的发展,越来越多的研究都强调了利用神经网络算法 NN (Neural Network) 监测地球大气和环境的优点和效率。神经网络算法目前已经被应用到不同卫星平台的气溶胶反演工作中,并取得了预期的效果。文中应用的是一个具有反向传播功能的多层感知 MLP (Multi-layer Perceptron) 神经元网络模型学习算法,MLP 是经典的机器学习算法之一,是利用不同类型像元光谱差异进行雾霾遥感识别的不二之选。MLP 属于前馈神经网络模型,其中至少包含一个隐含层,隐含层是作为一个通用函数逼近器存在。除此之外,MLP 结构中还包括输入层和输出层,每一个层级之间紧密相连。在有足够的隐藏神经元和训练数据的基础上,前馈神经网络能够学习输入与输出之间的数学关系。除输入层和输出层以外的层级都被称为隐含层。算法的主要工作流程如下:对于全部训练样本,先把样本输入到输入层神经元中,并逐层进行信号传递,直到得到输出结果;接下来计算输出层的误差,并且将计算出的误差依次反向传播给上一层神经元;最后根据每次传回的误差,调整连接权值和阈值。模型训练即为不停地重复迭代这个过程,直到训练次数达到设定的训练次数上限或者训练误差达到设定的训练误差为止。本文中神经网络模型训练的输入层即为 MODIS 可见光波段的 Rrc,隐含层为一层,每层包含 90 个隐藏神经元,通过不停地迭代训练,输出层即为对不同类型像元的识别结果。

2.3 机器学习算法流程

在模型训练前,建立一个三维向量,对三维向量的每一维分别设置空值和有效值,用于区分不同类型的像元,如用 (0,0,1) 表征清洁水体, (0,1,0) 表征浑浊水体, (1,0,0) 表征雾霾。对于 30% 的样本,可见光波段的光谱则为模型的输入,在隐含层的训练下,输出的结果即为表征不同像元的三维向量,即可以直接判别出每个像元的类型。模型建立好以后,剩余 70% 的数据将用于模型精度验证,即检验模型训练结果与真实像元结果的误差,本文则是通过识别率来表征模型精度。

具体工作流程见图 3,首先对 MODIS L1B 数据做校正,提取出 Rrc 信息;然后通过目视解译的方法,提取感兴趣区域,即不同类型的像元光谱信息;将 Rrc 作为输入,通过多层前馈神经网络模型的训练,识别出雾霾像元。

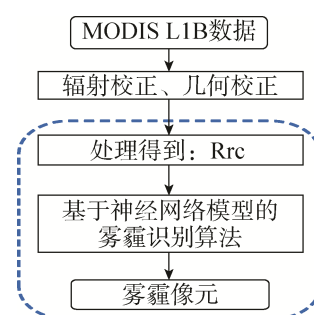


图 3 雾霾识别算法流程

Fig. 3 Flowchart of the NN model used for haze detection algorithm

3 结果与讨论

3.1 不同类型像元光谱特征分析

针对所选取的不同类型的样本,本文基于 Rrc 光谱差异实现对雾霾的遥感识别,文中对比分析了清洁水体、浑浊水体以及雾霾的光谱特征,以此为据展开对雾霾的遥感识别算法建立工作。不同类型的像元光谱特征如图 4~图 6 所示。图 4 给出了清洁水体的光谱曲线,图中灰黑色的线条代表不同样本的光谱,蓝线表示平均光谱的变化,蓝线上下两块浅蓝色区域表征的是光谱的标准差,从图中可以明显看出清洁

水体的光谱有随着波长增大而减小的趋势, 且清洁水体的 R_{rc} 值相对较低, 最大值也低于 0.07。

图 5 提供了浑浊水体的光谱特征分布。从数值上看, 浑浊水体的 R_{rc} 明显高于清洁水体的 R_{rc} , 最大值接近 0.16, 光谱随波长的变化趋势与清洁水体也不一样。波长小于 645 nm 时, R_{rc} 随波长的增大而增大, 且增幅较大; 波长大于 645 nm 时, R_{rc} 随波长的增大而减小, 且在 859 nm 波长之后, R_{rc} 的值基本处在低于 0.03 的位置。由此可见, 根据光谱的变化特征可以完全将清洁水体与浑浊水体区分开。

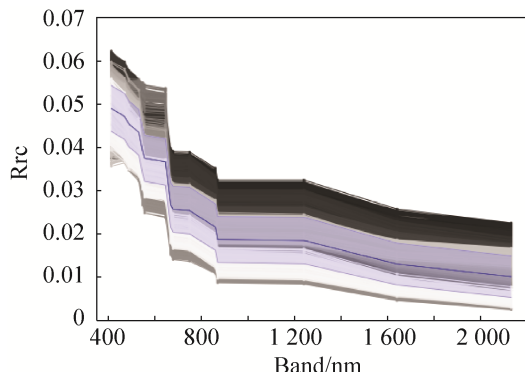


图 4 清洁水体的光谱特征分布

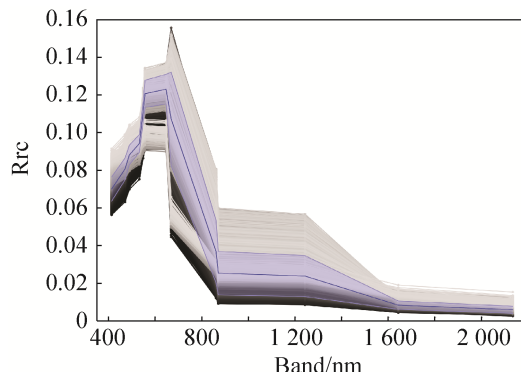


图 5 浑浊水体的光谱特征分布

Fig. 4 Rayleigh-corrected reflectance of clear water

Fig. 5 Rayleigh-corrected reflectance of turbid water

图 6 展示的是雾霾像元的光谱特征分布情况。从图中可以看到, 雾霾像元的光谱数值变化范围与浑浊水体近似, 其光谱随波长变化的趋势也与浑浊水体的 R_{rc} 随波长变化趋势类似。尽管如此, 但二者还是有所不同: 645 nm~667 nm 波长之间, 浑浊水体的 R_{rc} 随波长增大而减小, 而雾霾的 R_{rc} 随波长增加而增大; 雾霾的光谱随波长减小的趋势相对较为平缓; 且相对浑浊水体而言, 1 600 nm 波长之后的雾霾 R_{rc} 值相对较高。由此可知, 根据波长大于 645 nm 时 R_{rc} 的变化趋势以及 1 600 nm 波长之后的 R_{rc} 值可以区分开浑浊水体与雾霾。

综上所述, 根据不同类型像元的光谱差异可以将雾霾像元与其它像元区分开, 进而利用机器学习算法实现雾霾像元的遥感识别。

3.2 晴空条件下雾霾机器学习算法识别结果

本文借助 Python 语言, 建立雾霾遥感识别的神经网络机器学习算法。首先输入 30% 的训练样本, 对模型进行训练, 直到训练出最优模型, 再通过剩余 70% 的验证样本对模型的精度进行验证。算法反演结果如图 7 所示。黑色区域表征的是陆地, 紫色表征的是雾霾, 黄色表征的是浑浊水体, 绿色表征的是清洁水体。对比图 7 (左) 与图 2 (左), 从真彩色图像与反演算法的结果对比中可以发现, 机器学习算法反演出的雾霾空间分布情况与真彩图中目视解译的结果高度一致, 渤海海域上空的雾霾得到了较好的呈现, 表明该反演算法的精度较高。将该算法应用到一幅新图像, 即 2013 年 10 月 5 日的图像 (图 2 右) 中, 反演结果与真实情况也高度一致 (图 7 右), 识别准确度超过了 98%。这不仅验证了算法的反演精度, 另一方面也表明该算法适用于不同时间的晴空像元, 即验证了算法的适用性。由此可见, 通过光谱差异而建立的雾霾识别机器学习算法可以很好地提取雾霾像元, 适用于晴空条件下海上雾霾型气溶胶的识别与监测。

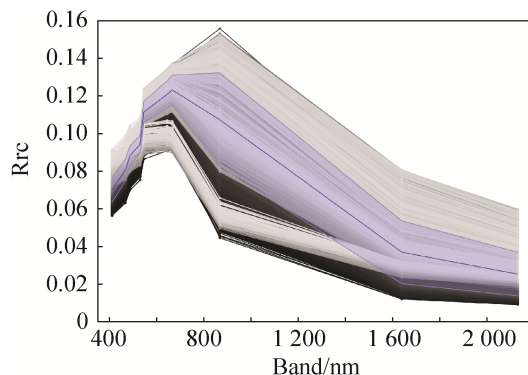


图 6 雾霾像元的光谱特征分布

Fig. 6 Rayleigh-corrected reflectance of haze pixel

值得注意的是, 本文所选取的神经网络模型为最基本最简单的模型, 样本选取时也假设不同类型像元的边界无明显差异。在未考虑复杂天气条件的情况下, 文中建立的雾霾遥感识别算法属于简单基础的模型。该算法尽管精度很高, 但由于缺少对边界情况、薄云等复杂条件的考虑, 算法只适用于晴空条件。

未来还要考虑更多的影响因素,对算法做改进,提高算法的适用性。

4 结束语

雾霾是较为典型的大气污染现象,是大气气溶胶污染的重要体现,对人类的生、活以及环境、气候等都产生了负面影响。近年来,雾霾已成为气象、环保部门以及科研工作者的重点关注现象。

因此,对雾霾的监测与预报显得尤为重要。传统的观测手段费时费力,且监测的空间尺度较小,遥感的监测手段具有时空覆盖范围广、连续性强等优势。本文设计了一种对中国近海上空雾霾的遥感识别与监测的机器学习算法。在晴空条件下,基于不同类型像元的光谱差异和神经网络方法,建立了雾霾遥感识别的机器学习算法,并对算法进行了验证。结果表明,该算法能够有效地识别出雾霾区,算法精度较高,识别准确度超过了 98%。并通过多景图像验证了算法的适用性,表明该算法完全适用于晴空条件下的雾霾识别。后期将针对算法存在的误差,展开相应的研究和改进工作。下一步可以尝试提高算法精度,建立一种适用于不同天气条件下的雾霾识别算法,以达到更方便快捷地识别雾霾的效果。

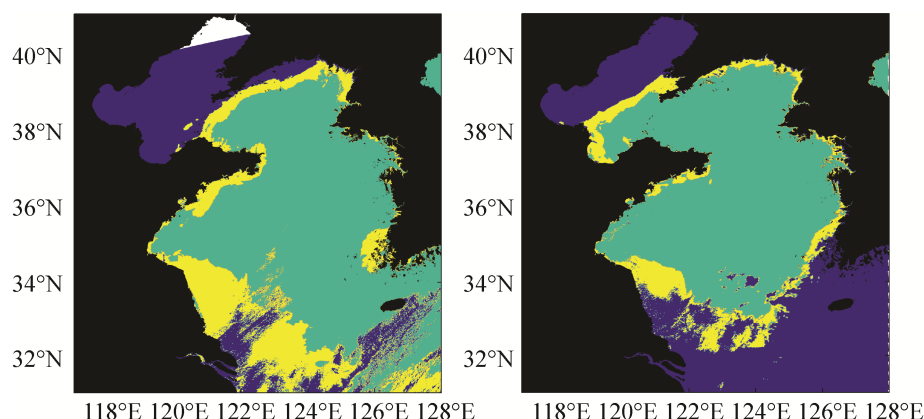


图 7 2013 年 10 月 4 日、5 日雾霾遥感监测结果

Fig. 7 Remote sensing detection of haze on October 4 & 5, 2013

参考文献

- [1] 张建立,全球海平面变化规律及比容变化的影响[D]. 青岛:中国海洋大学,2006.
- [2] HUSAR R B, PROSPERO J M, STOWE L L. Characterization of tropospheric aerosols over the oceans with the NOAA advanced very high resolution radiometer optical thickness operational product[J]. Journal of Geophysical Research Atmospheres, 1997, 102(D14): 16889–16909.
- [3] SIFAKIS N I, SOULAKELLIS N A. Satellite image processing for haze and aerosol mapping (SIPHA): code description and presentation of results[J]. IEEE International Geoscience & Remote Sensing Symposium, 2000, 1: 222–224.
- [4] 兰措,张永新,冬季西宁市上空阴霾的监测与分析[J]. 气象, 1988(6): 27–29.
- [5] ZHANG Y, GUINDON B, CIHLAR J. Development of a robust haze removal algorithm: assessment using temporally invariant targets[J]. Geoscience and Remote Sensing Symposium, 2002.
- [6] LIU J, SAITO Y, WANG H, et al. Stratigraphic development during the Late Pleistocene and Holocene offshore of the Yellow River delta, Bohai Sea[J]. Journal of Asian Earth Sciences, 2009. 36(4): 318–331.
- [7] GU Wei, SHI Peijun, LIU Yang, et al. The characteristics of temporal and spatial distribution of negative accumulated temperature in Bohai Sea and north Yellow Sea [J]. Journal of Natural Resources, 2002 (02): 168–173.
- [8] BARNES B B, HU C. Island building in the South China Sea: detection of turbidity plumes and artificial islands using Landsat and MODIS data[J]. Scientific reports, 2016, 6: 33194.
- [9] REMER L A, KAUFMAN Y J, TANRÉ D, et al. The MODIS aerosol algorithm, products, and validation[J]. Journal of the Atmospheric Sciences, 2005, 62: 947–973.
- [10] 刘阳, 陈丹, 贺克斌. 多角度成像光谱辐射仪在大气污染研究中的应用回顾[J]. 中国科学: D 辑, 2008. 38(3): 384–396.

[作者简介]

沈晓晶 1990 年生, 博士, 工程师, 主要研究方向为海洋遥感。

姜祝辉 1982 年生, 博士, 高级工程师, 主要研究方向为海洋遥感。

陈建 1986 年生, 博士, 高级工程师, 主要研究方向为海洋遥感。

刘娟 1978 年生, 博士, 高级工程师, 主要研究方向为海洋数值模拟。

马明 1977 年生, 硕士, 高级工程师, 主要研究方向为大气环境监测。