

## 光学学报

## 基于神经网络反演中国南海海域透明卷云参数

陆文强<sup>1,2,3</sup>, 杨世植<sup>1,2,3\*</sup>, 罗涛<sup>1,2,3</sup>, 李学彬<sup>1,2,3</sup>, 崔生成<sup>1,2,3</sup>, 程晨<sup>2,5</sup>, 韩露<sup>2,4</sup>, 石建军<sup>1,2,3</sup>, 韩叶颜<sup>1,2,3</sup><sup>1</sup>中国科学院合肥物质科学研究院安徽光学精密机械研究所中国科学院大气光学重点实验室, 安徽合肥 230031;<sup>2</sup>中国科学技术大学研究生院科学岛分院, 安徽合肥 230026;<sup>3</sup>先进激光技术安徽省实验室, 安徽合肥 230037;<sup>4</sup>中国科学院合肥物质科学研究院安徽光学精密机械研究所基础科学研究中心, 安徽合肥 230031;<sup>5</sup>中国科学院合肥物质科学研究院安徽光学精密机械研究所通用光学定标与表征技术重点实验室, 安徽合肥 230031

**摘要** 基于中分辨率成像光谱仪/云和气溶胶探测激光雷达(MODIS/CALIOP)的匹配数据集,提出一个分类神经网络进行透明卷云的识别,利用两个回归神经网络对透明卷云的光学厚度和云顶高度进行反演。结果表明,分类网络的精确度可达到84%,检测率达到79%。对成功识别的透明卷云参数进行了反演,得到透明卷云光学厚度的平均绝对误差为0.2,均方根误差为0.25,相关系数为0.79。对云顶高度进行了反演,得到云顶高度的平均绝对误差为0.61 km,均方根误差为0.74 km,相关系数达到0.87。本研究利用MODIS/CALIOP匹配数据集以及神经网络算法,可得到透明卷云的分布及其参数特性,为在南海海域上空的分布情况提供了数据支撑,有助于相关研究人员了解该地区透明卷云的分布情况,提高辐射计算的精度。

**关键词** 大气光学;透明卷云;中分辨率成像光谱仪;云和气溶胶探测激光雷达;神经网络;中国南海海域

中图分类号 P407.4

文献标志码 A

DOI: 10.3788/AOS230605

## 1 引言

卷云位于对流层上部和平流层下部,云顶气压小于440 hPa,多由冰晶粒子组成<sup>[1-2]</sup>。卷云分布广泛,约覆盖地球面积20%~35%,热带地区的卷云覆盖面积达到70%<sup>[3]</sup>。卷云通过反射太阳短波辐射和吸收地表热辐射,在地气系统中扮演着非常重要的角色<sup>[4]</sup>。其次,卷云会干扰空间信息的传输,对空间目标识别的准确性造成很大的影响<sup>[5]</sup>。此外,卷云是激光大气传输的重要影响因素,随着卷云光学厚度的增加,激光的透过率会大大降低<sup>[6]</sup>。张秀再等<sup>[7]</sup>的研究表明,卷云对量子通信光信号的传输有极大的影响。

卫星遥感探测卷云的手段分为被动探测和主动探测<sup>[8]</sup>。其中,被动探测器具有空间覆盖广、时间分辨率高等优势,主动探测器具有高灵敏度以及可提供目标的空间垂直结构信息的优点<sup>[9-10]</sup>。然而,被动探测器对薄卷云的探测有限,漏检了大量光学厚度小于0.4的薄卷云<sup>[11-12]</sup>,导致辐射计算产生较大的误差。Gao等<sup>[13]</sup>的研究表明,热带地区近40%的薄卷云被中分辨率成像光谱仪(MODIS)官方云检测算法标记为晴空像元。Sun等<sup>[14]</sup>指出光学厚度小于0.3的薄卷云可使日均反

射太阳短波辐射量增加 $2.5 \text{ W}\cdot\text{m}^{-2}$ ,光学厚度为0.1的薄卷云可使热辐射增加 $15 \text{ W}\cdot\text{m}^{-2}$ 。搭载在CALIPSO卫星上的双波长激光雷达CALIOP对薄卷云的探测具有无可比拟的优势,可有效探测光学厚度小于2的薄卷云,但由于消光作用,其对厚卷云的探测能力不足,而且空间覆盖范围有限<sup>[15-16]</sup>。

近年来,由于神经网络强大的非线性函数拟合能力,国内外已有许多研究将神经网络应用在了云参数反演领域。国内方面:2022年,Yang等<sup>[17]</sup>基于Himawari-8与CALIOP匹配数据集,利用机器学习方法进行了云检测和云顶高度反演,研究结果优于相关官方产品,但只限于光学厚度大于0.2的云。2019年,Min等<sup>[18]</sup>基于Himawari-8数据和CALIOP数据,利用机器学习方法对云顶高度进行了反演,研究结果证明了机器学习方法优于传统的物理方法。谭仲辉等<sup>[19]</sup>于2019年基于风云4A数据和CALIOP/Cloudsat数据,利用随机森林算法反演了云底高度,研究结果表明该算法对单层云的云底高度的估计结果较好,对多层云的估计结果较差。国外方面:2022年,Amell等<sup>[20]</sup>基于气象卫星Meteosat-9与Cloudsat匹配数据集,利用神经网络反演了冰水路径,反演结果与官方产品有较好

收稿日期: 2023-02-28; 修回日期: 2023-04-13; 录用日期: 2023-05-22; 网络首发日期: 2023-06-28

基金项目: 国家自然科学基金面上项目(41875041)、国家自然科学基金青年项目(42105082)、安徽省自然科学基金杰出青年项目(2008085J19)、先进激光技术安徽省实验室基金青年项目(AHL 2021QN01)

通信作者: \*szyang@aiofm.ac.cn

的一致性。2021年,McHardy等<sup>[21]</sup>基于静止环境卫星 GOES-16 与 CALIOP 匹配数据集,利用机器学习方法对海域透明卷云进行了检测,结果表明该方法可有效检测出 50% 的光学厚度小于 0.03 的不可见卷云,但该方法只限于白天使用。2014年,Kox等<sup>[9]</sup>基于 CALIOP 与静止轨道成像仪 SEVIRI 的匹配数据集,利用神经网络对卷云的光学厚度和云顶高度进行了反演;2017年,Strandgren等<sup>[22]</sup>在此基础上提出了一个分类网络,先对卷云进行识别后进行参数反演,进而提高了反演精度。上述研究的不足在于引入了多层云的影响,由于存在大量的透明卷云,因此,遥感器接收的辐射信息可能包含多层云的信息。本文主要对单层透明卷云进行识别和参数反演,不考虑多层云。

本文基于 2009—2015 年同在 A-Train 轨道上的 CALIOP 和 MODIS 数据,利用神经网络对中国南海海域上空的单层透明卷云进行识别和参数反演。

## 2 实验数据与方法

### 2.1 MODIS 和 CALIOP

搭载在 Aqua 卫星上的 MODIS 可获得从可见光 (0.405  $\mu\text{m}$ ) 到红外 (14.385  $\mu\text{m}$ ) 共计 36 个波段的观测数据。每 1~2 d 对地球观测一次,获取陆地和海洋温度、云、气溶胶等目标的图像<sup>[23]</sup>。本文针对中国南海海域 (0°N~25°N, 100°E~125°E), 收集了 2009—2015 年的 MYD02 数据,作为神经网络的输入特征参数。

CALIOP 由于敏感性较高,以及可以探测目标的垂直结构信息,它的探测数据经常和被动遥感数据结合起来,以提供更为全面的云层空间信息<sup>[24]</sup>。本文使用的是 CALIOP Level2\_05 km 云层产品。由于 MODIS MYD02 数据的空间分辨率为 1 km,因此需要对其进行重采样至 5 km,以便进行数据匹配。匹配方法采用最近邻法,数据处理流程如图 1 所示。

MODIS 数据转换公式如下:

$$\begin{cases} r = (M - d_{\text{offset}}) F_{\text{scale}} \\ B_1 = \frac{c_1}{\lambda \cdot \ln\left(1 + \frac{c_2}{r\lambda^5}\right)}, \end{cases} \quad (1)$$

式中:  $r$  为辐射值;  $M$  为 MODIS 原始数据;  $c_1 = 14390$ ;  $c_2 = 1.191 \times 10^8$ ;  $F_{\text{scale}}$  为校正因子;  $d_{\text{offset}}$  为偏置。MODIS 部分红外波段校正系数如表 1 所示。

读取的 CALIOP 数据参数有特征分类标记、云-气溶胶区分置信度 CAD\_Score、532 nm 特征光学厚度、特征层顶高度、特征层数以及经纬度。CALIOP 数据处理中的位运算是特征分类标记进行操作。该参数用 16 位二进制数表示,通过位运算可以得到透明卷云的信息。

本文根据 CAD\_Score、特征层数以及 MODIS/CALIOP 像元距离进行数据筛选。CAD\_Score 的

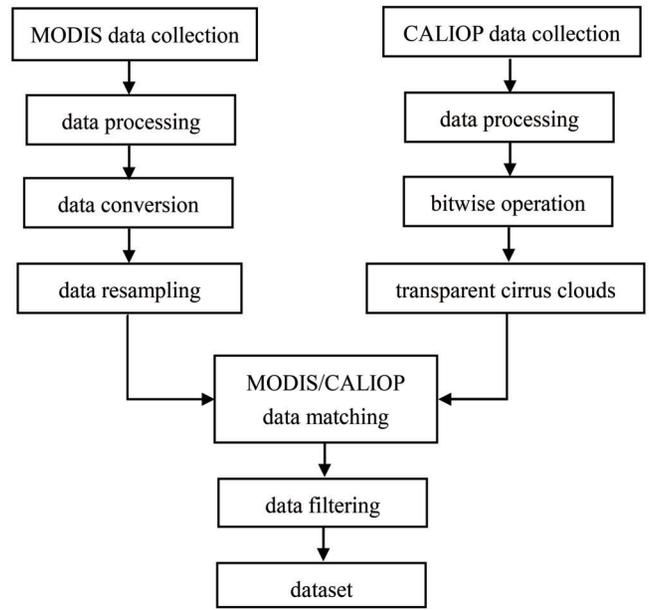


图 1 数据处理流程图

Fig. 1 Data processing flowchart

表 1 MODIS 部分红外波段校正系数

Table 1 MODIS correction coefficients for infrared bands

Band No.	$\lambda / \mu\text{m}$	$F_{\text{scale}} / 10^{-4}$	$d_{\text{offset}}$
27	6.715	1.178	2724.218
28	7.325	1.924	2317.488
29	8.550	5.571	2610.015
30	9.730	4.063	1560.333
31	11.030	6.508	2035.933
32	12.020	5.710	2119.085
33	13.335	2.623	2500.599
34	13.635	2.009	2499.094
35	13.935	1.768	2500.521
36	14.235	1.186	2495.891

介于 -100 和 100 之间,其中 0~100 表示云的分类置信度, -100~0 则代表气溶胶分类的置信度。CAD\_Score 的绝对值越大,分类正确的可信度就越高。本文只选取 CAD\_Score 大于 50 且特征层数为 1 的云层信息,同时保证 MODIS/CALIOP 之间的距离不超过 5 km。

根据上述三个条件进行筛选后,得到 2009—2015 年的数据样本共 274786 条,其中单层透明卷云样本 120980 条,晴空样本 153806 条。为提升透明卷云样本比例以提高模型的准确率,本文对晴空样本以 0.6 的系数进行随机降采样,得到 92284 个晴空样本,最终得到一个包含 213264 个样本的数据集。本文按 6:3:1 的比例将数据集分为训练集、测试集和验证集。

### 2.2 BP 神经网络

BP 神经网络应用广泛,它通过激活函数来实现非线性拟合。本文采用 Relu 函数和 Sigmoid 函数作为激

活函数。Relu 函数的数学表达式为

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (2)$$

式(2)表明,当 Relu 函数的输入小于 0 时,其输出为 0,表示不激活该神经元,从而稀疏化网络,以加快模型训练。

Sigmoid 函数的数学表达式为

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

式(3)表明,Sigmoid 函数将输入数据映射到(0, 1),且单调递增。本文仅在分类神经网络最后的输出层采用 Sigmoid 函数输出概率值。

分类问题中,二元交叉熵是很常见的损失函数,它描述了两种分布之间的差异。回归问题使用输出值与真值的均方误差作为损失函数。本文选用自适应矩估计 Adam 优化器。Adam 在训练过程中可以自适应地调整学习率,从而提高模型的收敛速度。

此外,本文使用早停法 Early Stopping 防止模型过拟合。实验中,训练批次 batch\_size 设为 128,使用 He 初始化方法对权重和偏置进行初始化。隐含层神经元的数量通常是输入特征个数的 1~2 倍,本文有 12 个输入参数,因此隐含层的神经元数量设为 16。

基于上述讨论,本文提出 3 个神经网络。一个分类神经网络用来识别透明卷云,使用 3 层隐含层,每层 16 个神经元,激活函数为 Relu 函数,输出层采用 Sigmoid 函数输出概率值,损失函数为二元交叉熵,选择 Adam 优化器;一个回归神经网络用来反演透明卷云的光学厚度(COD);一个回归神经网络用来反演透明卷云的云顶高度(CTH)。两个回归神经网络均采用 3 层隐含层、每层 16 个神经元的结构,激活函数为 Relu 函数,损失函数为均方误差 MSE,使用 Adam 优化器。

网络以 MODIS 各波段的亮温、亮温差等信息作为输入,包括 bt6.7  $\mu\text{m}$ 、bt7.3  $\mu\text{m}$ 、bt8.5  $\mu\text{m}$ 、bt11  $\mu\text{m}$ 、bt12  $\mu\text{m}$ 、bt13.34  $\mu\text{m}$ 、bt13.64  $\mu\text{m}$ 、bt13.94  $\mu\text{m}$ 、btd6.7-7.3、btd8.5-12、btd11-12。考虑到卷云随季节的变化情况,本文加入了当天天数的信息( $D$ ),且为了消除 2009—2015 年每年的 12 月 31 日至次年 1 月 1 日之间的差距,将其转换成正弦值  $\sin(2\pi \cdot D/365)$ ,将 CALIOP 数据作为标签值。在训练之前,需要对输入数据进行标准化,以缩小数据量级之间的差距。标准化公式如下:

$$x_i = \frac{x_i - \bar{x}}{\sigma} \quad (4)$$

式中: $x_i$ 是第  $i$  个样本; $\bar{x}$ 是样本均值; $\sigma$ 是样本标准差。

### 2.3 验证指标

本文选取 7 个指标对结果进行分析,其中,检测率  $P_{\text{OD}}$  和误报率  $R_{\text{FA}}$  用于对透明卷云识别结果的分析;平

均绝对误差  $M_{\text{AE}}$ 、平均误差百分比  $M_{\text{PE}}$ 、平均绝对误差百分比  $M_{\text{APE}}$ 、均方根误差  $R_{\text{MSE}}$  和相关系数  $R$  用于对反演结果的分析。各指标的表达式如下:

$$P_{\text{OD}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (5)$$

$$R_{\text{FA}} = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TN}}} \quad (6)$$

式中: $N_{\text{TP}}$ 表示正确分类为透明卷云的样本数量; $N_{\text{FN}}$ 表示未被正确分类的透明卷云的样本数量; $N_{\text{FP}}$ 表示被错误分类为透明卷云的样本数量; $N_{\text{TN}}$ 表示被正确分类的晴空的样本数量。由该 4 个量组成的矩阵称为混淆矩阵。

$$M_{\text{AE}} = \frac{1}{N} \sum_{i=1}^N |P_i - T_i| \quad (7)$$

$$M_{\text{PE}} = \frac{1}{N} \sum_{i=1}^N \frac{P_i - T_i}{T_i} \times 100\% \quad (8)$$

$$M_{\text{APE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_i - T_i}{T_i} \right| \times 100\% \quad (9)$$

$$R_{\text{MSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - T_i)^2} \quad (10)$$

$$R = \frac{\sum_{i=1}^N (P_i - \bar{P})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2 (T_i - \bar{T})^2}} \quad (11)$$

式中: $P_i$ 和  $T_i$ 分别表示第  $i$  个样本的预测值和真值。

## 3 分析与讨论

### 3.1 透明卷云参数的分布

图 2 展示了 CALIOP 探测的 2009—2015 年中国南海海域单层透明卷云的光学厚度和云顶高度的统计信息。

图 2(a)表明,透明卷云的光学厚度主要分布在 1.5 以内。其中,光学厚度在 0~0.1 之间的透明卷云约为 36%,光学厚度大于 2 的透明卷云占比非常小,因此,本文只针对光学厚度小于 2 的透明卷云进行反演。图 2(b)展示了透明卷云的云顶高度分布。云顶高度主要分布在 12~17 km 以内,其中在云顶高度为 16 km 处出现峰值,约为 19%。经过透明卷云的识别之后,参数的各范围比例发生了明显的变化。其中,在 0~0.1 内的透明卷云的光学厚度从 36% 下降到了 23% 左右,说明该部分透明卷云由于光学厚度太小而难以与晴空区分。云顶高度的分布未出现明显的变化。本文针对数据分布不平衡的问题,对光学厚度大于 1.5 的部分透明卷云进行了 1 倍过采样。

### 3.2 透明卷云的识别结果分析

图 3(a)、(b)分别给出了检测率  $P_{\text{OD}}$  和误报率  $R_{\text{FA}}$  的接受者操作特性曲线(ROC)和检测率随光学厚度的变化曲线。对于分类神经网络的输出概率值,不同

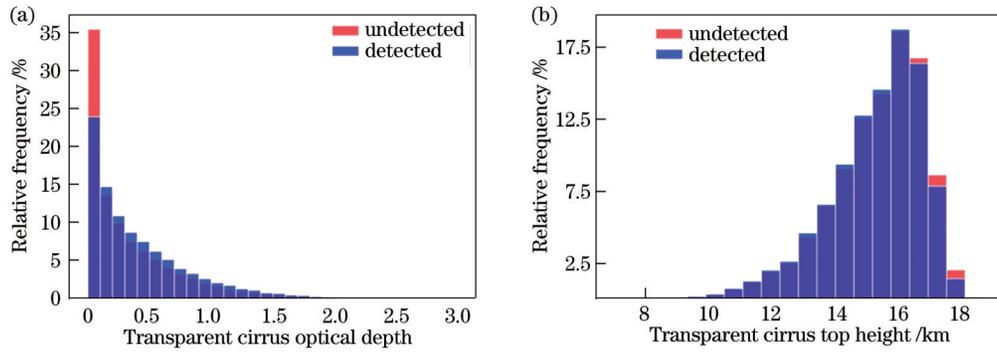


图2 透明卷云光学厚度和云顶高度。(a)透明卷云光学厚度;(b)云顶高度

Fig. 2 Histograms of transparent cirrus cloud optical depth and top height. (a) Transparent cirrus cloud optical depth; (b) top height of cirrus cloud

的阈值对应不同的检测率和误报率。图3(a)中的圆点为最佳阈值对应的检测率 $P_{OD}$ 和误报率 $R_{FA}$ 。确定

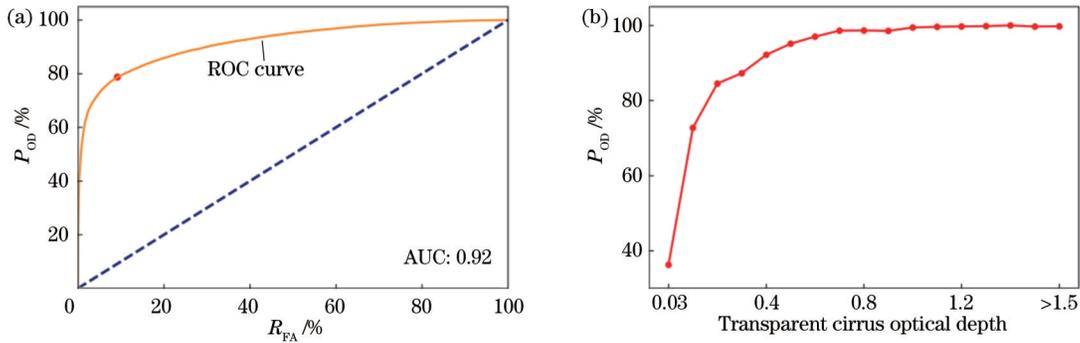


图3 检测率 $P_{OD}$ 与误报率 $R_{FA}$ 的ROC曲线以及检测率 $P_{OD}$ 随透明卷云光学厚度的变化曲线。(a) ROC曲线;(b)检测率 $P_{OD}$ 随透明卷云光学厚度的变化曲线

Fig. 3 ROC curve of probability of detection  $P_{OD}$  and rate of false alarm  $P_{OD}$ , and probability of detection  $P_{OD}$  varying with transparent cirrus optical depth. (a) ROC curve; (b) probability of detection  $P_{OD}$  varying with transparent cirrus optical depth

在识别透明卷云时,将网络输出的概率阈值设为0.56可达到最优的检测率和误报率,检测率为79%,误报率为9.8%,AUC达到0.92。检测率较低和误报率较高是因为存在大量光学厚度小于0.1的透明卷云。基于该阈值,得到对应的混淆矩阵(表2)。

图3(b)表明,检测率随着光学厚度的增大而迅速上升。当光学厚度小于0.03时,检测率仅为36%;当光学厚度在0.03~0.1之间时,检测率迅速上升到73%;当透明卷云的光学厚度大于0.4时,检测率均达到95%以上。因此,该神经网络可以有效地检测出透明卷云,但对于光学厚度小于0.03的卷云,其表现较差。

表2 混淆矩阵

Table 2 Confusion matrix

Item		Predicted value	
		Transparent cirrus clouds	Clear sky
True value	Transparent cirrus clouds	28873 ( $N_{TP}$ )	7509 ( $N_{FN}$ )
	Clear sky	2716 ( $N_{FP}$ )	24881 ( $N_{TN}$ )

最佳阈值后,设置不同的光学厚度区间,可得到检测率与光学厚度之间的关系,如图3(b)所示。

### 3.3 透明卷云参数反演结果分析

图4给出了透明卷云光学厚度和云顶高度的平均误差百分比 $M_{PE}$ 和平均绝对误差百分比 $M_{APE}$ 的分布情况。从图4(a)和4(c)中发现,光学厚度的误差大多分布在100%以内,约占75%。其中,误差在50%以内的约占53%。图4(b)和4(d)表明云顶高度的平均误差百分比分布较为平衡,呈现高斯分布,误差主要分布在10%以内。

图5展示了平均误差百分比 $M_{PE}$ 和平均绝对误差百分比 $M_{APE}$ 随参数值的变化曲线。图5(a)表明,当光学厚度小于0.03时,其误差较大,这是因为该部分卷云的红外信号太弱。当光学厚度逐渐变大时,误差迅速下降至100%以内。当光学厚度大于0.2时,误差均在50%以内。导致光学厚度整体误差较大的原因是神经网络对光学厚度小于0.03的卷云的反演结果误差较大。图5(b)展示了误差随云顶高度的变化情况。当云顶高度在15.5~16 km时, $M_{PE}$ 接近于0, $M_{APE}$ 主要分布在5%左右。可以看出,在云顶高度小于10 km和大于17.5 km的部分,误差较大,主要原因是这两个区间的样本量较少,代表性较差。

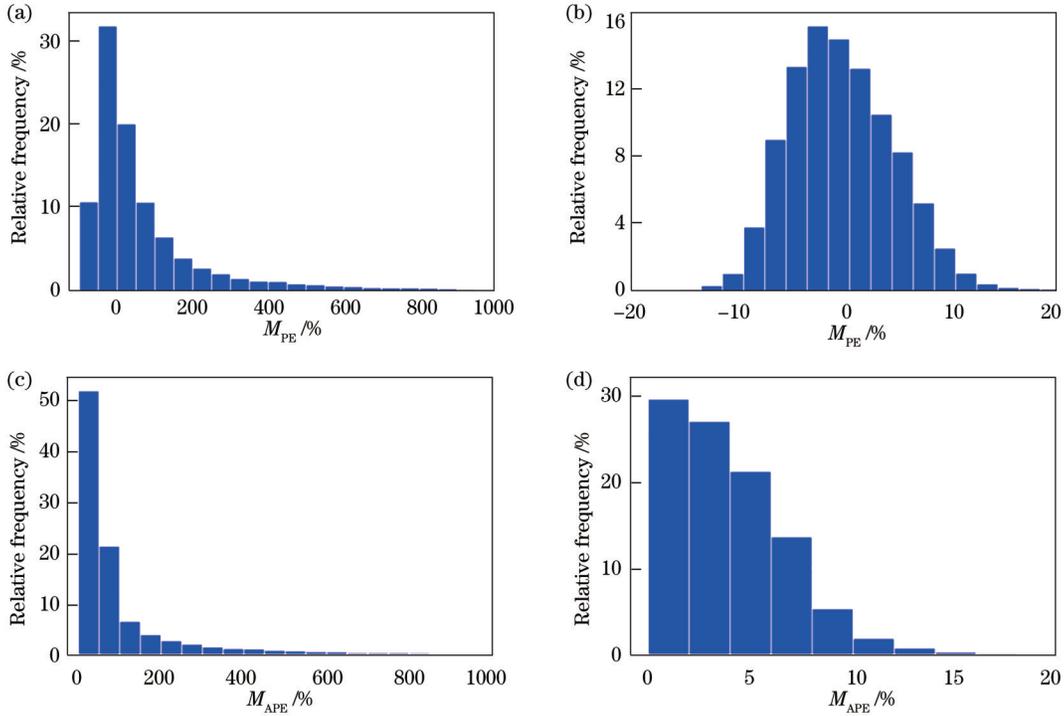


图 4 透明卷云光学厚度和云顶高度的  $M_{PE}$  和  $M_{APE}$ 。(a) 光学厚度的  $M_{PE}$ ; (b) 云顶高度的  $M_{PE}$ ; (c) 光学厚度的  $M_{APE}$ ; (d) 云顶高度的  $M_{APE}$ 。  
Fig. 4  $M_{PE}$  and  $M_{APE}$  of optical depth and top height of transparent cirrus clouds. (a)  $M_{PE}$  of optical depth; (b)  $M_{PE}$  of top height; (c)  $M_{APE}$  of optical depth; (d)  $M_{APE}$  of top height

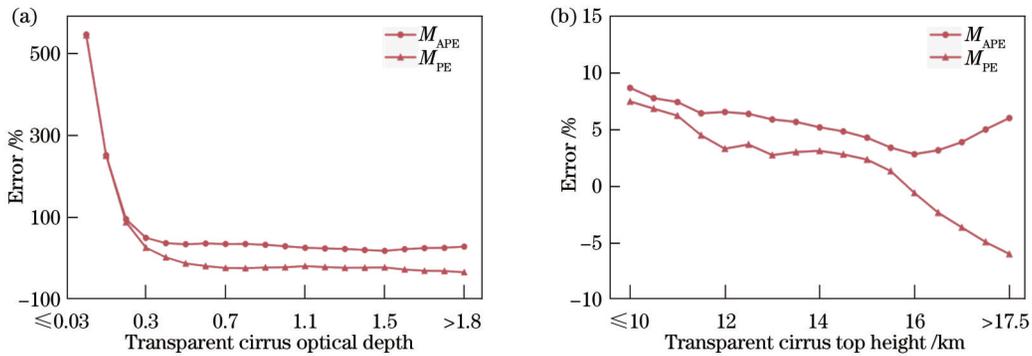


图 5  $M_{PE}$  和  $M_{APE}$  随光学厚度和云顶高度的变化曲线。(a) 光学厚度; (b) 云顶高度  
Fig. 5  $M_{PE}$  and  $M_{APE}$  varying with optical depth and top height. (a) Optical depth; (b) top height

图 6(a)、(b) 分别展示了光学厚度和云顶高度的预测值和真值的散点图,同时给出了预测值和真值之间的相关系数、均方根误差以及平均绝对误差。

图 6 表明,预测值和真值存在较好的线性关系,相关系数分别达到 0.79 和 0.87。图 6(a) 显示:当光学厚度小于 0.5 时,神经网络模型存在高估的情况;当光学厚度大于 0.5 时,神经网络模型存在低估的情况。图 6(b) 表明,云顶高度的预测值和真值的分布较为均匀,未出现明显的高估和低估情况。表 3 给出了均方根误差  $R_{MSE}$  和相关系数  $R$  与相关研究对比的结果。其中,均方根误差  $R_{MSE}$  与对比数据没有明显的差别,而光学厚度的相关系数  $R$  均大于相关研究结果,原因是本文仅对单层透明卷云进行研究,而相关研究引入了多层

云的影响。

### 3.4 个例分析

图 7 是 2008 年 5 月 26 日 MODIS 经过南海海域时观测到的云量信息,其中直线是 CALIOP 的轨迹。

对任何云量值大于 0 的像元,本文都认为是云像元。经过阈值处理以及透明卷云神经网络的检测得到图 8,其中 1 表示云,0 表示晴空。

图 8 表明,经网络检测后的云量明显多于 MODIS 官方产品的云量。为了验证该检测效果,将网络部分检测结果与 CALIOP 的观测数据进行了对比分析(图 9)。首先将 MODIS 与 CALIOP 进行匹配,得到的 MODIS 晴空观测结果对应图 9(a)。神经网络检测结果对应图 9(b)。图 9(c) 给出了 CALIOP 的探测信息。

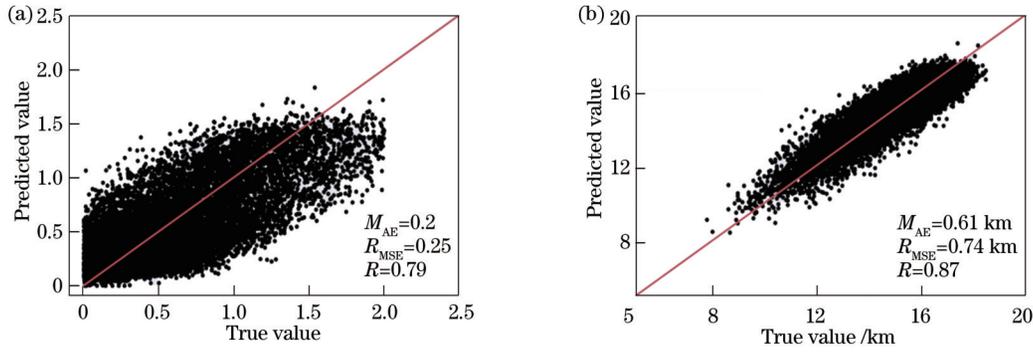


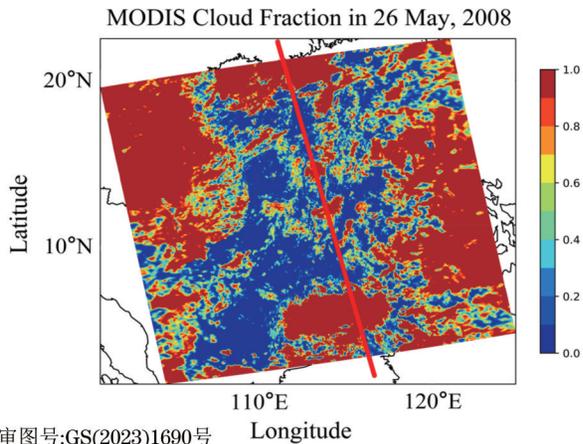
图 6 预测值与真值的散点图。(a)光学厚度;(b)云顶高度

Fig. 6 Scatter plots of predicted values and true values. (a) Optical depth; (b) top height

表 3 算法对比结果

Table 3 Comparison of results of related algorithms

Algorithm	$R_{MSE}/R$	
	Optical depth	Top height
Algorithm proposed by Kox <i>et al.</i>	0.24/0.61	0.71 km/0.82
Algorithm proposed by Strandgren <i>et al.</i>	Not given/0.65	Not given/0.90
Ours	0.25/0.79	0.74 km/0.87



审图号:GS(2023)1690号

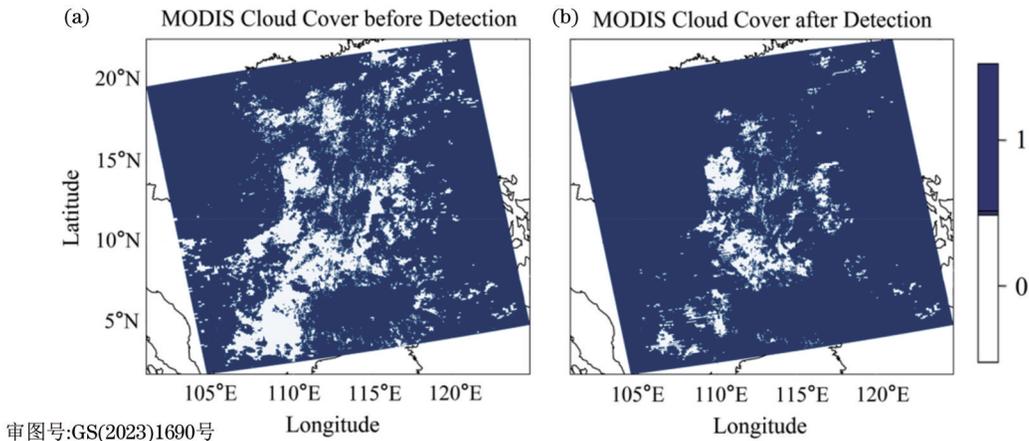
图 7 MODIS 云量和 CALIOP 轨道(直线)

Fig. 7 MODIS cloud fraction and CALIOP trajectory (line)

可以看到,相比于 MODIS 官方产品,检测之后的结果更加符合 CALIOP 的探测结果。通过将这些检测出来的透明卷云输入到神经网络中,反演其光学厚度与云顶高度信息,再与 CALIOP 真值数据进行对比(图 10)。

图 10(a)展示了光学厚度的反演结果与 CALIOP 真值的对比情况。可以看到,对于光学厚度小于 0.1 的透明卷云,反演值与真值的误差较大。对于光学厚度大于 0.1 的透明卷云,反演值与真值较一致。图 10 (b)是云顶高度反演值与真值的对比情况。可以看到,反演值与真值之间的一致性较好。与 CALIOP 数据的对比验证表明,本文结果具有一定的可靠性。

图 11 给出了 MODIS 漏检的透明卷云。图 12(a)和(b)分别展示了光学厚度和云顶高度分布。从图 12 中可以看到,MODIS 漏检的这些云的光学厚度较小,



审图号:GS(2023)1690号

图 8 MODIS 产品云量分布。(a)检测前;(b)检测后

Fig. 8 Distributions of MODIS cloud fraction. (a) Before detection; (b) after detection

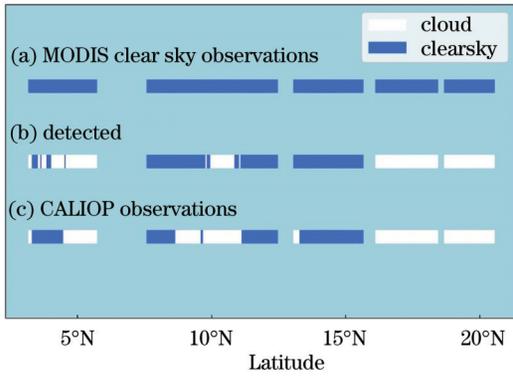


图 9 检测结果对比。(a)MODIS 晴空观测结果;(b)神经网络检测结果;(c)CALIOP 观测结果  
 Fig. 9 Comparison of detection results. (a) MODIS clear sky observation results; (b) detected results of neural network; (c) CALIOP observation results

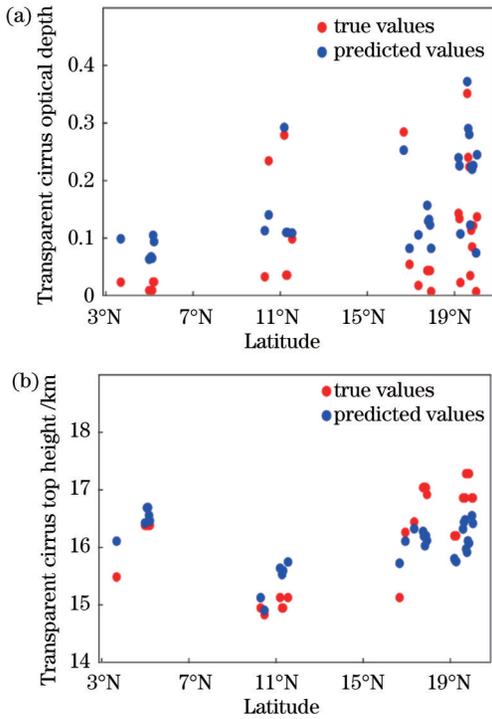


图 10 反演值与真值对比。(a)光学厚度;(b)云顶高度  
 Fig. 10 Comparison of inverse values and true values. (a) Optical depth; (b) top height

大部分在 0.3 以内,且云顶高度较高,主要分布在 14 km 以上。

### 4 结 论

本文基于 MODIS 与 CALIOP 的匹配数据集,利用神经网络首先对透明卷云进行识别检测,再对成功识别出的透明卷云进行光学厚度和云顶高度的反演,并将结果与 CALIOP 真值数据进行比较。结果显示,分类神经网络能够较好地检测出透明卷云,检测率为 79%。反演结果与真值具有较好的一致性。其中:光学厚度的相关性为 0.79,优于相关研究;云顶高度的

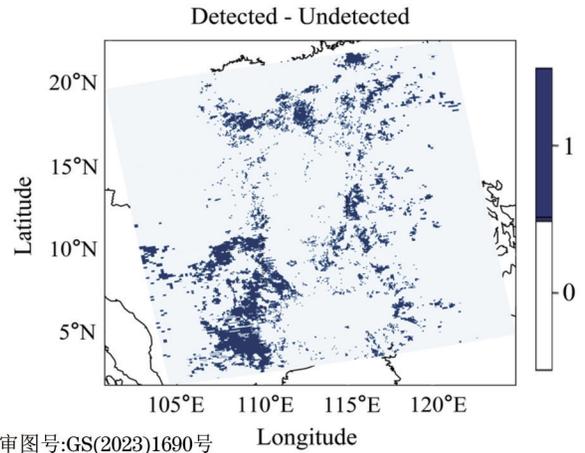


图 11 MODIS 漏检的部分透明卷云  
 Fig. 11 Transparent cirrus clouds undetected by MODIS

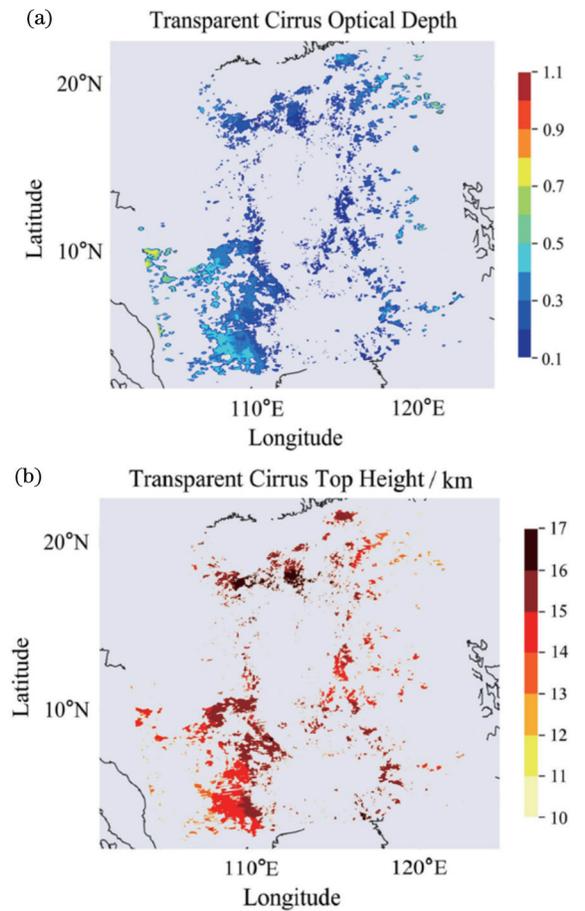


图 12 MODIS 漏检透明卷云的光学厚度和云顶高度分布。  
 (a)光学厚度;(b)云顶高度  
 Fig. 12 Distributions of optical depth and top height of transparent cirrus clouds undetected by MODIS. (a) Optical depth; (b) top height

相关性达到 0.87。神经网络模型对光学厚度小于 0.03 的不可视卷云的检测能力较差,检测率只有 36%,且反演误差较大。随着光学厚度的增大,检测率迅速上升。当光学厚度大于 0.4 时,检测率达到 95%

以上,且反演误差迅速下降。最后对实例进行验证分析。结果表明,相比于MODIS官方产品,神经网络的检测结果更加符合CALIOP的观测结果;之后对识别出来的透明卷云进行参数反演,反演结果与真值存在较好的一致性。研究结果为MODIS漏检的透明卷云的分布情况提供了一定的参考价值,有利于提高辐射计算的精度。

### 参 考 文 献

- [1] Cziczo D J, Froyd K D, Hoose C, et al. Clarifying the dominant sources and mechanisms of cirrus cloud formation[J]. *Science*, 2013, 340(6138): 1320-1324.
- [2] Liou K N. Influence of cirrus clouds on weather and climate processes: a global perspective[J]. *Monthly Weather Review*, 1986, 114(6): 1167-1199.
- [3] Dessler A E, Yang P. The distribution of tropical thin cirrus clouds inferred from Terra MODIS data[J]. *Journal of Climate*, 2003, 16(8): 1241-1247.
- [4] Sassen K, Campbell J R. A midlatitude cirrus cloud climatology from the facility for atmospheric remote sensing. part I: macrophysical and synoptic properties[J]. *Journal of the Atmospheric Sciences*, 2001, 58(5): 481-496.
- [5] 蔡熠, 刘延利, 戴聪明, 等. 卷云大气条件下目标与背景对比度模拟分析[J]. *光学学报*, 2017, 37(8): 0801001.  
Cai Y, Liu Y L, Dai C M, et al. Simulation analysis of target and background contrast in condition of cirrus atmosphere[J]. *Acta Optica Sinica*, 2017, 37(8): 0801001.
- [6] Ren S H, Gao M, Wang M G, et al. Attenuation and transmission characteristics of laser propagation in cirrus clouds with a spherical boundary[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(1): 316-321.
- [7] 张秀再, 徐茜, 刘邦宇. 卷云对自由空间星地量子通信信道的影响[J]. *光学学报*, 2021, 41(20): 2027001.  
Zhang X Z, Xu X, Liu B Y. Influence of cirrus clouds on space-to-earth quantum communication channels in free space[J]. *Acta Optica Sinica*, 2021, 41(20): 2027001.
- [8] 叶松, 付思梦, 李树, 等. 基于MODIS参数的卷云光学厚度计算方法[J]. *激光与光电子学进展*, 2021, 58(19): 1901003.  
Ye S, Fu S M, Li S, et al. Optical thickness calculation method of cirrus based on MODIS parameters[J]. *Laser & Optoelectronics Progress*, 2021, 58(19): 1901003.
- [9] Kox S, Bugliaro L, Ostler A. Retrieval of cirrus cloud optical thickness and top altitude from geostationary remote sensing[J]. *Atmospheric Measurement Techniques*, 2014, 7(10): 3233-3246.
- [10] 季承荔, 陶宗明, 胡顺星, 等. 三波长激光雷达探测合肥地区卷云特性[J]. *光学学报*, 2014, 34(4): 0401001.  
Ji C L, Tao Z M, Hu S X, et al. Cirrus measurement using three-wavelength lidar in Hefei[J]. *Acta Optica Sinica*, 2014, 34(4): 0401001.
- [11] Ackerman S A, Holz R E, Frey R, et al. Cloud detection with MODIS: part II. Validation[J]. *Journal of Atmospheric and Oceanic Technology*, 2008, 25(7): 1073-1086.
- [12] Meyer K, Platnick S. Utilizing the MODIS 1.38  $\mu\text{m}$  channel for cirrus cloud optical thickness retrievals: algorithm and retrieval uncertainties[J]. *Journal of Geophysical Research: Atmospheres*, 2010, 115(D24): 209-222.
- [13] Gao B C, Kaufman Y J, Tanre D, et al. Distinguishing tropospheric aerosols from thin cirrus clouds for improved aerosol retrievals using the ratio of 1.38- $\mu\text{m}$  and 1.24- $\mu\text{m}$  channels[J]. *Geophysical Research Letters*, 2002, 29(18): 1890-1891.
- [14] Sun W B, Videen G, Kato S, et al. A study of subvisual clouds and their radiation effect with a synergy of CERES, MODIS, CALIPSO, and AIRS data[J]. *Journal of Geophysical Research: Atmospheres*, 2011, 116(D22): 207-217.
- [15] Mao F Y, Pan Z X, Henderson D S, et al. Vertically resolved physical and radiative response of ice clouds to aerosols during the Indian summer monsoon season[J]. *Remote Sensing of Environment*, 2018, 216: 171-182.
- [16] Young S A, Vaughan M A, et al. CALIOP algorithm theoretical basis document[EB/OL]. [2023-02-28]. [https://xueshu.baidu.com/usercenter/paper/show?paperid=5c9bbe8f2169f252729ac1312647eb27&site=xueshu\\_se](https://xueshu.baidu.com/usercenter/paper/show?paperid=5c9bbe8f2169f252729ac1312647eb27&site=xueshu_se).
- [17] Yang Y K, Sun W X, Chi Y L, et al. Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data[J]. *Remote Sensing of Environment*, 2022, 273: 112971.
- [18] Min M, Li J, Wang F, et al. Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms[J]. *Remote Sensing of Environment*, 2020, 239: 111616.
- [19] 谭仲辉, 马烁, 韩丁, 等. 基于随机森林算法的FY-4A云底高度估计方法[J]. *红外与毫米波学报*, 2019, 38(3): 381-388.  
Tan Z H, Ma S, Han D, et al. Estimation of cloud base height for FY-4A satellite based on random forest algorithm[J]. *Journal of Infrared and Millimeter Waves*, 2019, 38(3): 381-388.
- [20] Amell A, Eriksson P, Pfreundschuh S. Ice water path retrievals from Meteosat-9 using quantile regression neural networks[J]. *Atmospheric Measurement Techniques*, 2022, 15(19): 5701-5717.
- [21] McHardy T M, Campbell J R, Peterson D A, et al. Advancing maritime transparent cirrus detection using the advanced baseline imager "cirrus" band[J]. *Journal of Atmospheric and Oceanic Technology*, 2021, 38(6): 1093-1110.
- [22] Strandgren J, Bugliaro L, Sehnke F, et al. Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks[J]. *Atmospheric Measurement Techniques*, 2017, 10(9): 3547-3573.
- [23] Barnes W L, Xiong X, Salomonson V V. Status of terra MODIS and aqua modis[J]. *Advances in Space Research*, 2003, 32(11): 2099-2106.
- [24] Sassen K, Wang Z E, Liu D. Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) measurements[J]. *Journal of Geophysical Research*, 2008, 113: D00A12.

# Parameter Retrieval of Transparent Cirrus Clouds over South China Sea Based on Artificial Neural Networks

Lu Wenqiang<sup>1,2,3</sup>, Yang Shizhi<sup>1,2,3\*</sup>, Luo Tao<sup>1,2,3</sup>, Li Xuebin<sup>1,2,3</sup>, Cui Shengcheng<sup>1,2,3</sup>,  
Cheng Chen<sup>2,5</sup>, Han Lu<sup>2,4</sup>, Shi Jianjun<sup>1,2,3</sup>, Han Yeyan<sup>1,2,3</sup>

<sup>1</sup>Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, Anhui, China;

<sup>2</sup>Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, Anhui, China;

<sup>3</sup>Advanced Laser Technology Laboratory of Anhui Province, Hefei 230037, Anhui, China;

<sup>4</sup>Center for Fundamental Science Research, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, Anhui, China;

<sup>5</sup>Key Laboratory of Optical Calibration and Characterization, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, Anhui, China

## Abstract

**Objective** Cirrus clouds are located in the upper troposphere and lower stratosphere and are mostly composed of ice crystal particles with a cloud top pressure of less than 440 hPa. Meanwhile, they are widely distributed, covering an area of approximately 20%–35% of the earth. In tropical regions, the coverage area of cirrus clouds reaches 70%. Cirrus clouds play a very important role in the earth atmosphere system by reflecting solar shortwave radiation and absorbing surface thermal radiation. Passive remote sensing features wide spatial coverage and high temporal resolution, but its detection for thin cirrus clouds is relatively limited. A large number of thin cirrus clouds with optical depth of less than 0.4 are missed by passive remote sensing. This phenomenon is particularly serious in tropical areas, which greatly underestimates the coverage of cirrus clouds and causes large errors in radiation calculation. Fortunately, the dual-wavelength cloud aerosol lidar infrared with orthogonal polarization (CALIOP) onboard the CALIPSO has unparalleled advantages in detecting thin cirrus clouds, which can detect thin cirrus clouds with optical depth less than 2. However, CALIOP cannot penetrate thick cirrus clouds, with limited spatial coverage. Therefore, we combine active and passive remote sensing to identify and retrieve optical depths and top heights of single-layer transparent cirrus clouds over the South China Sea. We hope that our research can provide quantitative support for the distribution of thin cirrus clouds over the South China Sea, and help improve the radiation calculation accuracy.

**Methods** Artificial neural networks have powerful nonlinear function fitting ability and can employ multi-channel information, which makes them widely applicable in cloud parameter retrieval. Based on a matched dataset from MODIS/CALIOP, we utilize back propagation (BP) neural networks for the identification and parameter retrieval of single-layer transparent cirrus clouds over the South China Sea. First, to obtain a perfect dataset, we conduct several steps, including MODIS/CALIOP data collection, data processing, data conversion, data resampling, data matching, and data filtering. Additionally, we collect MODIS/MYD02\_1 km and CALIOP/Level2\_CloudLayer\_5 km data for a total of six years from 2009 to 2015 over the South China Sea. MODIS raw data are converted to brightness temperature as inputs of neural networks. Due to the spatial resolution of 1 km for MYD02 data, it is necessary to resample it to 5 km for data matching with CALIOP. Meanwhile, data filtering based on three conditions is conducted, including CAD\_Score and Number\_Layers\_Found from CALIOP, and MODIS/CALIOP pixel distance. We set the threshold of CAD\_Score as 50. Since we only focus on single-layer transparent cirrus clouds, Number\_Layer\_Found is set as 1 and the distance between MODIS/CALIOP is ensured to be no more than one pixel (5 km). After filtering, a total of 274786 data samples were obtained from 2009 to 2015, including 120980 single-layer transparent cirrus samples and 153806 clear sky samples. To increase the proportion of positive samples (transparent cirrus clouds) to improve the model accuracy, we randomly downsample the clear sky samples by 0.6 times to obtain 92284 clear sky samples. Therefore, a dataset containing 213264 samples is finally acquired. We divide the dataset into training, testing, and validation sets in a 6 : 3 : 1 ratio, and three neural networks are proposed, including one for detecting transparent cirrus clouds, one for retrieving optical depths, and one for retrieving top heights.

**Results and Discussions** In the identification of transparent cirrus clouds, the probability threshold of the network output is set to 0.56, which can achieve the optimal detection rate and false alarm rate, with a detection rate of 79%, a false

alarm rate of 9.8%, and an AUC of 0.92 [Fig. 3(a)]. A large number of transparent cirrus clouds with optical depth less than 0.1 result in a low detection rate and a high false alarm rate, which is because it is difficult to distinguish these cirrus clouds from clear skies. The detection rate rapidly increases with the rising optical depth [Fig. 3(b)]. When the optical depth is less than 0.03 or greater than 0.4, the detection rate is only 36% or over 95% respectively. Among them, when the optical depth is greater than 1, the detection rate can reach 100%, indicating that the neural network can detect transparent cirrus clouds in the region, but the detection performance on the subvisual cirrus clouds with optical depth less than 0.03 is poor. For the optical depth retrieval of transparent cirrus clouds, the error exceeds 500% under the optical depth of less than 0.03. As the optical depth gradually increases, the error rapidly decreases to within 100%. When the optical depth is greater than 0.2, the error is within 50% [Fig. 5(a)]. The reason for the large overall error is that the neural network has a significant error in retrieving cirrus clouds with optical depth less than 0.03. The error for the top height of cirrus clouds is mainly distributed around 5%. In the parts with cloud top heights less than 10 km and greater than 17.5 km, greater errors can be observed [Fig. 5(b)], and the main reason for speculation is that the sample sizes in these two intervals are relatively small. The scatter plots reveal good linear relations between the predicted and true values (Fig. 6), with correlation coefficients reaching 0.79 (for optical depth) and 0.87 (for top height). The mean absolute error and root mean square error for the optical depth are 0.2 and 0.25 respectively. The mean absolute error and root mean square error of the top height retrieval are 0.61 km and 0.74 km respectively. In a case study, the comparison with CALIOP data shows that the research results have a certain reliability degree (Figs. 9 and 10).

**Conclusions** Based on the MODIS/CALIOP matched dataset, neural networks are adopted to first detect transparent cirrus clouds, then retrieve the optical depth and cloud top height of the detected transparent cirrus clouds, and compare the results with the CALIOP data. The results show that the classification neural network can detect transparent cirrus clouds with a detection rate of 79%. The retrieval results show that it has a high agreement with CALIOP data. The correlation of the optical depth is 0.79 and that of the cloud top height is 0.87. The classification neural network has poor detection ability for subvisual cirrus clouds with optical depth less than 0.03, with a detection rate of only 36% and a larger retrieval error. As the optical depth increases, the detection rate rises rapidly. For the parts with optical thickness greater than 0.4, the detection rate reaches more than 95% and the retrieval error decreases rapidly. Finally, we perform a case study. The results show that the detection results of the neural network are more consistent with the observations of CALIOP than the official MODIS product. The retrieval results are in good agreement with CALIOP data. The results can provide references for the distribution of transparent cirrus clouds missed by MODIS, and help improve the radiation calculation accuracy.

**Key words** atmospheric optics; transparent cirrus clouds; moderate-resolution imaging spectroradiometer; cloud aerosol lidar infrared with orthogonal polarization; neural networks; South China Sea