

# 基于随机森林算法的南海上空大气 CO<sub>2</sub> 柱浓度估算模型构建及其检验与应用\*

周芳成<sup>1,2</sup> 刘少军<sup>3,4,5</sup> 田光辉<sup>3,4,5</sup> 韩秀珍<sup>1,2</sup> 甘业星<sup>3,4,5</sup>

1 国家卫星气象中心（国家空间天气监测预警中心），北京 100081

2 许健民气象卫星创新中心，北京 100081

3 海南省气象科学研究所/海南省南海气象防灾减灾重点实验室，海口 570203

4 中国气象局三沙海洋气象野外科学试验基地，三沙 573199

5 海南省南海海洋气象野外科学观测研究站，三沙 573199

**摘要：**南海作为边缘海，其上空大气 CO<sub>2</sub> 含量受大陆、海洋、大气多重因素影响。本文利用多源宽幅卫星的叶绿素 a 浓度、瞬时光和有效辐射、颗粒无机碳、颗粒有机碳、海表温度、风速、风向 7 个参数，基于随机森林模型建立了南海大气 CO<sub>2</sub> 柱浓度估算模型，以 2020 年数据对模型精度验证，Bias 为 0.27 ppm，R<sup>2</sup> 为 0.59，RMSE 为 1.00 ppm，整体精度较高。研究发现，南海大气 CO<sub>2</sub> 柱浓度呈现明显的季节特征，表现为春季 > 夏季 > 冬季 > 秋季。造成南海大气 CO<sub>2</sub> 柱浓度季节差异的主要影响因素是随时间变化的，在 1 月和 4 月，主要影响因素是风向，7 月份风速和风向是影响最大的 2 个因素，10 月份海温成为最主要影响因素。本方法基于宽幅多源遥感数据，可实现对南海大气 CO<sub>2</sub> 柱浓度的高频次、全覆盖监测，。

**关键词：**南海；随机森林算法；大气 CO<sub>2</sub> 柱浓度；估算模型；季节变化特征

## Construction, test and application of atmospheric CO<sub>2</sub> column concentration estimation model over the South China Sea based on random forest model

ZHOU Fangcheng<sup>1,2</sup> LIU Shaojun<sup>3,4,5</sup> TIAN Guanghui<sup>3,4,5</sup> HAN Xiuzhen<sup>1,2</sup> Gan Yexing<sup>3,4,5</sup>

1 National Satellite Meteorological Center (National Center for Space Weather), Beijing 100081

2 Innovation Center for Fengyun Meteorological Satellite (FYSIC), Beijing 100081

3 Hainan Institute of Meteorological Science/ Key Laboratory of South China Sea Meteorological Disaster Prevention and Mitigation of Hainan Province, Haikou 570203

4 Sansha Marine Meteorology Field Experiment Station of China Meteorological Administration, Sansha 573199

5 South China Sea Marine Meteorology Hainan Observation and Research Station, Sansha 573199

**Abstract:** Common carbon satellites have narrow width and long revisit period, so it is difficult to achieve high frequency and full coverage CO<sub>2</sub> monitoring. The atmospheric CO<sub>2</sub> content over the South China Sea is affected by continental, oceanic and atmospheric factors. In this paper, a random-forest-based model of atmospheric CO<sub>2</sub> column concentration over the South China Sea was built with chlorophyll-a concentration, instantaneous photosynthetically active radiation, particulate inorganic carbon, particulate organic carbon, sea surface temperature, wind speed and wind direction, which were from multisource satellites remote sensing data. The accuracy of the model was verified for the year 2020, with Bias of 0.27 ppm, R<sup>2</sup> of 0.59 and the RMSE is 1.00 ppm, showing the accuracy is satisfactory. The results show that the atmospheric CO<sub>2</sub> column

\*基金支持：国家自然科学基金青年基金：有云条件下基于被动微波遥感的地表温度与发射率联合反演方法研究（42001309），中国气象局风云卫星应用先行计划（2022）(FY-APP-2022.0308)，海南省气象局科研项目（HNQXJS202214）

第一作者：周芳成，博士，高级工程师，主要从事生态气象应用研究。E-mail: [zhoufc@cma.gov.cn](mailto:zhoufc@cma.gov.cn)

通讯作者：刘少军，博士，正高级工程师，研究方向为卫星遥感应用。E-mail: [cdutlsj@163.com](mailto:cdutlsj@163.com)

concentration in the South China Sea presents obvious seasonal characteristics, which are spring>summer>winter>autumn. Moreover, the main influencing factors causing the seasonal differences of atmospheric CO<sub>2</sub> column concentration in the South China Sea vary with time. In January and April, the main influencing factors are wind direction. In July, wind speed and wind direction are the two most influential factors. In October, sea surface temperature is the most influential factor. This method can realize the high-frequency and full-coverage monitoring of atmospheric CO<sub>2</sub> column concentration in the South China Sea.

**Key words:** the South China Sea; random forest algorithm; atmospheric CO<sub>2</sub> column concentration; estimation model; seasonal variation characteristics

工业革命以来,化石燃料大量使用、森林砍伐、城市建设等因素已经深刻影响了地球的碳循环过程,大气中 CO<sub>2</sub> 浓度不断上升,温室效应加剧、极端天气事件频发(Marland et al, 1985; Tian et al, 2016; 乔年等, 2023)。为了积极应对气候变化,推进生态优先、节约集约、绿色低碳发展,我国提出了“碳达峰、碳中和”目标。在“双碳”目标的实现过程中,准确掌握大气中 CO<sub>2</sub> 含量及其变化规律具有重要意义。

大气中 CO<sub>2</sub> 含量的探测可分为地基和空基两种方法。地基探测方法具有较高的精度,如全球碳柱总量观测网(Total Carbon Column Observing Network, TCCON)(Wunch et al, 2011, 2015),是一个由地基傅立叶变换光谱仪组建的观测网络,可进行 CO<sub>2</sub>、CH<sub>4</sub> 等要素的精密测量,但是其缺点是站点数非常少,在全球范围内稀疏分布,无法反映大区域连续大气内 CO<sub>2</sub> 的分布趋势。空基探测一般是基于某些专用于大气温室气体探测的卫星,如日本的 GOSAT 和 GOSAT2、美国的 OCO-2 和 OCO-3、中国的碳卫星、高分五号卫星等(叶函函等, 2021; Crisp et al, 2017; Hakkarainen et al, 2016; Schneising et al, 2011; 邓安健等, 2020; 白文广等, 2010; 葛钰洁和肖钟湧, 2022),利用其携带的短波红外波段对大气 CO<sub>2</sub> 进行探测,其最大的优点是随着卫星的绕地运动具有了全球覆盖的探测能力,但是其通道特性决定了其幅宽窄、重访周期长,无法实现每日的全球覆盖探测。

温室气体探测卫星对大气 CO<sub>2</sub> 含量的探测多采用物理反演方法,利用辐射传输方程模拟整个光学路径,求解过程相对复杂(张炳炎等, 2021)。大气 CO<sub>2</sub> 含量除了受人类活动影响外,还与下垫面生物活动(光合作用、呼吸作用)有着密切关系,而生物活动又受到环境的温度湿度等因素影响(加亦琪等, 2022)。因此,大气 CO<sub>2</sub> 含量与下垫面生物种类、丰度、环境温湿度等因素有直接关系(He et al, 2017)。基于此,不少学者提出,通过对大气 CO<sub>2</sub> 含量的影响要素筛选,建立从影响要素到大气 CO<sub>2</sub> 含量的回归关系,从而直接估算大气 CO<sub>2</sub> 含量。如 Guo et al(2012)利用中分辨率成像光谱仪(Moderate Resolution Imaging Spectroradiometer, MODIS)的地表温度(Land Surface Temperature, LST)、归一化植被指数(Normalized Difference Vegetation Index, NDVI)、叶面积指数(Leaf Area Index, LAI)、净初级生产力(Net Primary Productivity, NPP)、总初级生产力(Gross Primary Productivity, GPP)等产品,建立与 GOSAT 探测的大气 CO<sub>2</sub> 柱浓度(the column-averaged CO<sub>2</sub> dry air mole fraction, XCO<sub>2</sub>)的多元线性回归方程,分别在全球建立不同的回归方程,精度为-2.56~3.14 ppm。后续 Guo et al(2015)将该方法进一步按照不同月份建立回归关系,应用于东亚的植被生长季节(5-10月)。也有学者利用神经网络、随机森林(Random Forest, RF)等机器学习方法,建立了新型的回归模型。如 Li et al(2022)选择植被覆盖度(Fractional Vegetation Cover, FVC)、LAI、NPP、蒸散(Evapotranspiration, ET)、温度(Temperature)、相对湿度(Relative Humidity, RH)、风向和风速(Wind Direction and Wind Speed)共 8 个要素,利用极限树(Extreme Random Tree)方法建立与大气 CO<sub>2</sub> 柱浓度回归关系。He et al(2023)利用高程(Digital Elevation Model, DEM)、人口密度(Population Density, PD)、土地利用(Land Use, LU)、NDVI、气象再分析数据(ECMWF Reanalysis v5, ERA5)数据,讨论了多种机器学习方法(随机森林、极限树等)对大气 CO<sub>2</sub> 柱浓度的预测精度,结果显示随机森林方法精度最高。李静波等(2023)利用 NDVI、ERA5、DEM、

气溶胶光学厚度(Aerosol Optical Depth, AOD)和地表反射率(Surface Reflectance, SR)数据,采用决策树和集成学习方法对大气 CO<sub>2</sub> 柱浓度进行预测,准确率超过 90%。

当前对 CO<sub>2</sub> 柱浓度的关注重点在陆地上空(蔡大鑫等, 2023),对海洋上空关注的较少。海洋是一个巨大的碳库,人类每年产生的人为 CO<sub>2</sub> 约有一半停留在大气中,而其余的 CO<sub>2</sub> 被海洋所吸收(Manabe et al, 1993; 詹滨秋和赵永平, 1989; 张远辉等, 2000),对海上大气 CO<sub>2</sub> 含量的研究,对于预测全球大气 CO<sub>2</sub> 含量乃至气候变化具有重要意义。海洋上空的 CO<sub>2</sub> 柱浓度受到海—气多种因素共同影响,边缘海上空的 CO<sub>2</sub> 柱浓度还会受到陆地因素影响,属于多要素复杂非线性问题,面对这样的问题,以随机森林为代表的机器学习方法表现出了与传统统计回归方法相比更高的精度。本文以中国南海上空的 CO<sub>2</sub> 柱浓度为研究对象,假设 OCO-2 的大气 CO<sub>2</sub> 柱浓度为真实值,筛选与 OCO-2 同时过境的多源卫星遥感大气和海洋参数,利用随机森林模型建立大气和海洋参数与大气 CO<sub>2</sub> 柱浓度的回归关系,在此基础上估算海洋上空大范围连续大气 CO<sub>2</sub> 柱浓度,并开展大气 CO<sub>2</sub> 柱浓度季节变化特征分析。

## 1 研究区与数据源

### 1.1 研究区概况

南海是我国面积最大的边缘海,也是我国最深、最大的海,总面积达 350 万 km<sup>2</sup>,南海常年盛行季风,冬季受强大的东北季风影响,夏季则受西南季风影响,并受到由季风控制的环流影响。作为边缘海,南海海上大气 CO<sub>2</sub> 柱浓度的时空变化除了受海—气 CO<sub>2</sub> 分压差影响外,还受其他因素影响,如陆地大气 CO<sub>2</sub> 浓度、风向风速、海表温度、水中浮游植物数量、水产养殖种类和规模等,对照这些影响因素,本文选择了相应的遥感数据(详见 1.3, 1.4)。

### 1.2 OCO-2 数据

美国于 2014 年 7 月发射 OCO-2 卫星,轨道高度 705 km,幅宽 10.6 km,升轨过境时间为地方时 13:36,空间分辨率 2.2 km×1.29 km,时间分辨率 16 天(Liang et al, 2017)。OCO-2 是当前国际主要的碳监测卫星之一,与 TCCON 对比精度优于 1ppm (Liang et al, 2017)。OCO-2 共有 3 个谱段,分别是 O<sub>2</sub>A 吸收的 0.76 μm、弱 CO<sub>2</sub> 吸收的 1.61 μm 和强 CO<sub>2</sub> 吸收的 2.06 μm,根据这 3 个谱段可反演大气 CO<sub>2</sub> 柱浓度。本文使用经过偏差校正的 2020 年 L2 级 Lite 产品,为了保证数据的可靠性,参与随机森林训练和验证的数据均为质量标志 QF=0 (质量为优)(Liang et al, 2017)。为了使其与其他遥感数据的空间分辨率一致,使用三次样条插值法将其空间分辨率重采样至 0.05°×0.05°(为确保真值可靠性,仅对有价值像元分辨率进行重采样,无值像元仍为无值)。

### 1.3 先进散射计(Advanced Scatterometer, ASCAT)数据

欧洲气象卫星应用组织(European Organisation for the Exploitation of Meteorological Satellites, EUMETSAT)分别于 2006、2012、2018 年发射了 MetOp-A、B、C 三颗极轨卫星,轨道高度 817 km,其上搭载的先进散射计(Advanced Scatterometer, ASCAT)是一台拥有 3 个垂直极化天线的 C 波段(5.2 GHz)微波散射计。ASCAT 可被用于测量海面风场(等效 10 m)、极地冰、土壤水分等。海面风场是通过微波散射计获取后向散射系数来探测海洋表面的粗糙度,再根据海洋表面粗糙度反演海洋表面的风速和风向,田光辉等(2020)用南海浮标站对风速风向产品进行了精度检验,其风速平均绝对偏差为 0.59 m/s,风向平均绝对偏差为 16.23°。本文使用的是 MetOp-C 星 2020 年 0.25° 空间分辨率的网格化日产品,选取海表面风速和风向 2 个参数。为了与其他遥感数据的空间分辨率一致,使用三次样条插值法将其空间分辨率重采样至 0.05°×0.05°。MetOp-C 星的过境时间为地方时 9:30,与 OCO-2 升轨过境时相差约 4 h,本文假设大气 CO<sub>2</sub> 柱浓度在这段时间内的变化可忽略。

### 1.4 MODIS 产品

海表温度日产品(Sea Surface Temperature, SST),白天,空间分辨率 4 km,全球海洋覆

盖。SST 能显著影响海洋生物的新陈代谢，进而影响浮游植物的呼吸和光合作用。

叶绿素 a 浓度日产品(Chlorophyll-a Concentration, Chl-a)，白天，空间分辨率 4 km，全球海洋覆盖。叶绿素 a 浓度的高低与水体中浮游植物的种类、数量以及水环境质量等高度相关，可以在一定程度上反映水质状况(孙军等, 2003)。

瞬时光合有效辐射日产品(Instantaneous Photosynthetically Active Radiation, IPAR)，白天，空间分辨率 4 km，全球海洋覆盖。浮游植物吸收太阳辐射中可见光波段的光谱进行光合作用，该波段的太阳光被称为光合有效辐射。植物通过叶绿素将这些光谱能量转化为植物生长所需的生物化学能量，此过程会吸收 CO<sub>2</sub> 并造成水中 CO<sub>2</sub> 分压变化。

颗粒有机碳日产品(Particulate Organic Carbon, POC)，白天，空间分辨率 4 km，全球海洋覆盖。海水吸收的 CO<sub>2</sub> 通过真光层内浮游植物的光合作用被转化为有生命的颗粒有机碳，表层海洋的 POC 沉降作为海洋生物泵的重要组成部分，提供了一种深海固碳机制。真光层底颗粒有机碳通量能够可反映深海固碳效率。

颗粒无机碳日产品(Particulate Inorganic Carbon, PIC)，白天，空间分辨率 4 km，全球海洋覆盖。PIC 被认为是表征海水中颗石藻产量的指标，在海洋碳循环过程中，颗石藻是主要的初级生产者和碳酸盐生产者。现代开放大洋中，约一半的碳酸盐沉积来自颗石藻的贡献，在海洋碳循环中扮演着重要的角色。

以上参数下载自美国 NASA 网站(<https://oceancolor.gsfc.nasa.gov>)，数据时间为 2020 年，使用三次样条插值法将 MODIS 产品重采样至 0.05°×0.05°。MODIS/AQUA 卫星升轨的过境时间为地方时 13:30，与 OCO-2 升轨过境时相差约 6 min，在此假设大气 CO<sub>2</sub> 柱浓度在这段时间内没有明显变化。

## 2 研究方法

### 2.1 随机森林模型

RF 模型是一种基于决策树集成的机器学习算法，有着对多要素、大数据量、复杂非线性问题求解的优势，可用于解决复杂的分类和回归问题(Breiman, 2001)。在对陆上大气 CO<sub>2</sub> 柱浓度的预测中，RF 方法相较于其他机器学习方法，显示了更高的精度(He et al, 2023)。RF 中的每棵树都通过随机且有放回的方法(bootstrap sample)在原始的训练集中抽取 2/3 训练实例，根据生成的样本集构建决策树，所有单个决策树的结果通过聚合得到最终的预测结果。对于每棵树，未被抽取到训练集中的 1/3 的数据叫做袋外数据(Out Of Bag, OOB)，可以用来评估所构建的模型精度，即 OOB 特征重要性。

### 2.2 模型验证

随机森林模型对海上大气 CO<sub>2</sub> 柱浓度的估算精度由 OCO-2 卫星的 XCO<sub>2</sub> 作为真值进行检验，具体指标为决定系数(R<sup>2</sup>)、均方根误差(Root Mean Square Error, RMSE)和偏差(Bias)。

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

$$Bias = \frac{\sum_{i=1}^n (Y_i - X_i)}{n}$$

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

其中，X<sub>i</sub> 为 OCO-2 卫星监测到的 XCO<sub>2</sub>，Y<sub>i</sub> 为经过随机森林模型估算的 XCO<sub>2</sub>， $\bar{X}$  为 OCO-2 卫星监测到的 XCO<sub>2</sub> 的平均值，单位均为 ppm。

## 3 结果与讨论

### 3.1 随机森林模型设置

随机森林模型需要设置叶子和树的数目，为确定两者的最优数目，分别设置叶子数为 5, 10, 20, 50, 100, 200, 500，树的数目为 1-200，计算模型的均方误差(Mean Squared Error,

MSE)。结果如图 1 所示，当叶子数目为 5 或 10 时 MSE 的线最低，本文取最优的叶子数为 5；当树的数目达到 80 时 MSE 处于最小并基本稳定，即最优的树的数目是 80。

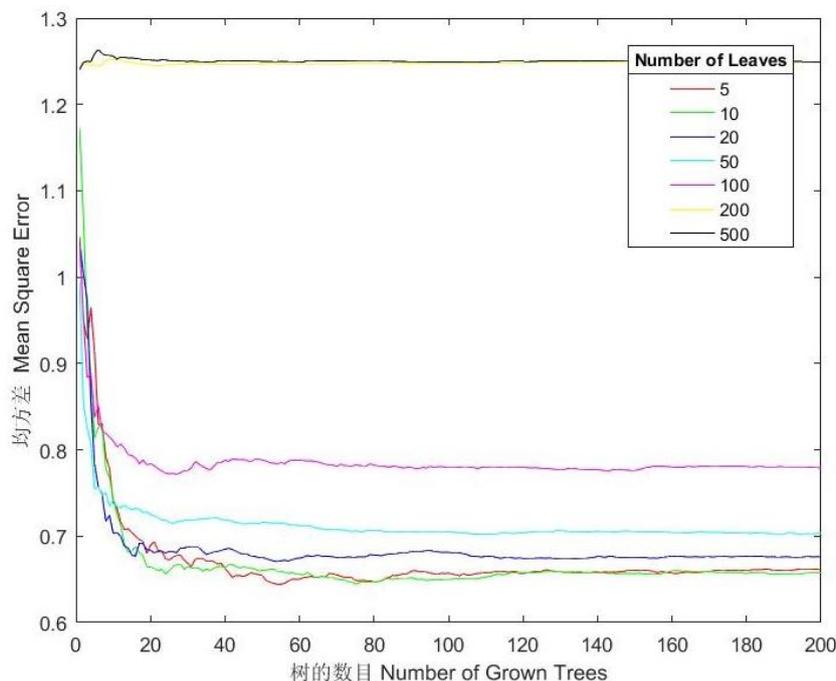


图 1 模型误差随叶子和树的数目变化关系

Fig. 1 Model error varies with the number of leaves and trees

### 3.2 模型精度验证

随机选取 2020 年 4/5 的数据作为随机森林模型的训练数据集，1/5 作为验证数据集，总数据量 1774，符合随机森林方法对样本数量的需求，验证结果如图 2 所示。如图可知，验证数据集的实际大气 CO<sub>2</sub> 柱浓度变化范围约为 407.4~414.7 ppm，随机森林模型的估算范围为 410~413.6 ppm，模型表现出 0.27 ppm 的高估，高估主要出现在 CO<sub>2</sub> 柱浓度低值区，而在高值区则表现出一定的低估，大气 CO<sub>2</sub> 柱浓度在 413 ppm 附近时模型的精度相对较高。模型的 R<sup>2</sup> 为 0.59，RMSE 为 1.00 ppm，较前人陆地上空 CO<sub>2</sub> 柱浓度估算精度(R<sup>2</sup> 为 0.00~0.59，RMSE 为 1.17~1.79 ppm)更高(Guo et al, 2015)，这也与南海上空 CO<sub>2</sub> 柱浓度年内变化幅度弱于陆地上空有一定关系。

分别选取 1、4、7、10 月作为冬季、春季、夏季、秋季的代表月份对模型精度进行验证，结果见表 1。根据表中结果，模型在春夏两季精度较高，R<sup>2</sup> 最高达到 0.65，RMSE 最小达到 0.70 ppm；在秋冬两季精度略差，R<sup>2</sup> 最低为 0.49，RMSE 最大达到 1.24 ppm。从单月南海大气 CO<sub>2</sub> 柱浓度平均值的对比来看，也是春夏两季精度较高，相差均为 0.05 ppm，秋冬两季分别相差 0.07 ppm 和 0.10 ppm。

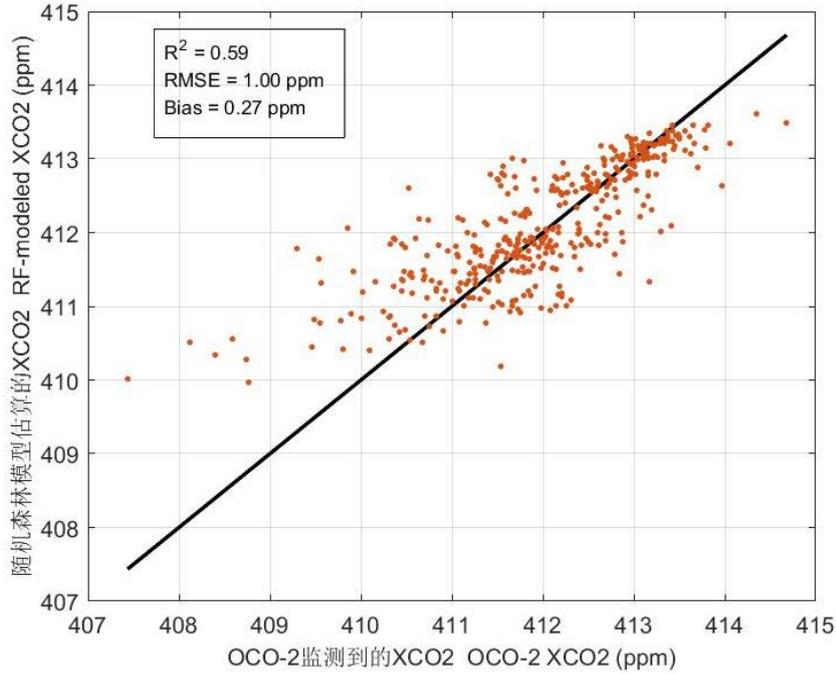


图2 随机森林模型精度验证

Fig. 2 Scatterplot for accuracy verification of random forest models

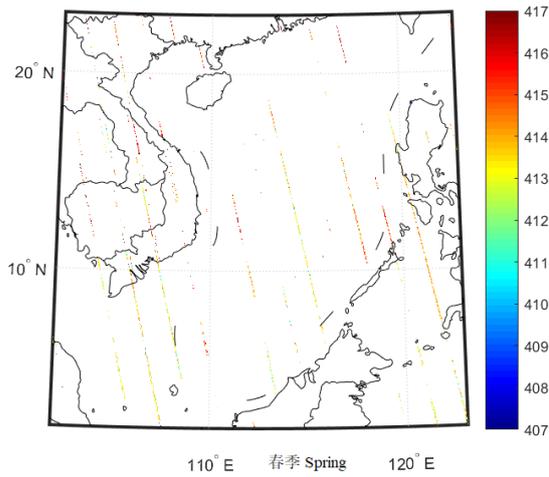
表1 不同季节精度验证

Tab. 1 Accuracy verification table in different seasons

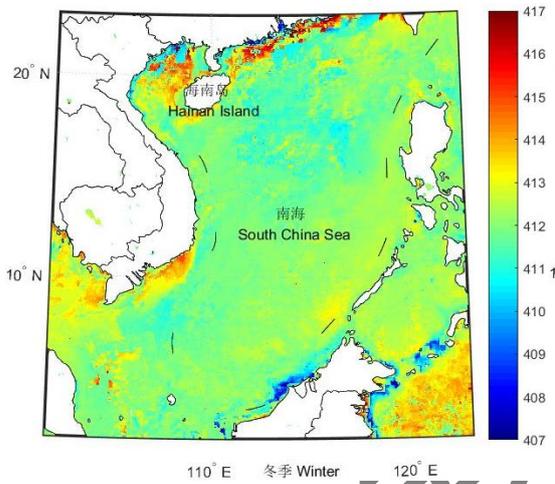
指标	冬季	春季	夏季	秋季
R <sup>2</sup>	0.49	0.65	0.57	0.50
RMSE (ppm)	1.19	0.83	0.70	1.24
南海平均值—OCO-2(ppm)	411.81	413.83	412.56	410.90
南海平均值—模型(ppm)	411.91	413.78	412.61	410.83

### 3.3 南海大气 CO<sub>2</sub> 柱浓度分布季节特征

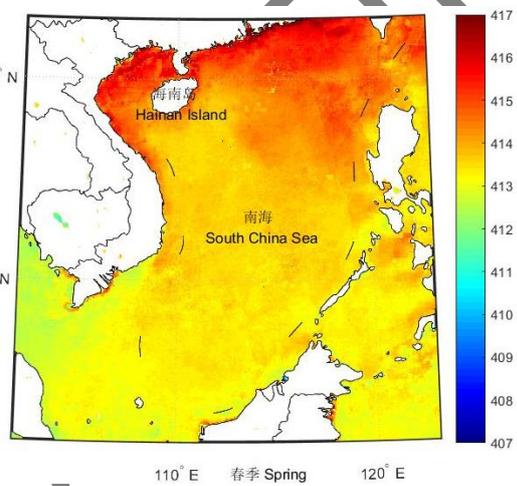
基于宽幅多源卫星构建的南海大气 CO<sub>2</sub> 柱浓度随机森林模型具有高重访周期的优势，以此构建的月或季平均大气 CO<sub>2</sub> 柱浓度数据基本可实现对南海区域的空间全覆盖，相较于 OCO-2 卫星的窄幅数据，能够更好的反映南海大气 CO<sub>2</sub> 柱浓度的空间分布特征(见图 3)。图 3a 展示了 OCO-2 卫星在 2020 年 4 月份在南海上空大气 CO<sub>2</sub> 柱浓度的分布，图 3b、3c、3d、3e 分别展示了基于宽幅多源卫星数据的南海区域在 2020 年的大气 CO<sub>2</sub> 柱浓度随季节变化的分布特征，具体表现为南海大气 CO<sub>2</sub> 柱浓度春季>夏季>冬季>秋季。如图 3b、3c、3d、3e 所示，春季南海上空大气 CO<sub>2</sub> 柱浓度整体较高，南海北部海域特别是近海岸带附近高值达到 415~417 ppm，南海西部和南部的大气 CO<sub>2</sub> 柱浓度相对略低，低值区为 412~413 ppm；夏季大气 CO<sub>2</sub> 柱浓度有所降低，高值区在北部湾、海南岛西海岸、南海东北部，低值区为南海西部、南部和东南部；秋季大气 CO<sub>2</sub> 柱浓度整体在全年属于最低，海盆区域高于近海岸附近，南海北部特别是近海岸带附近、南部海岸带附近值最低，低值达到 407~410 ppm；冬季大气 CO<sub>2</sub> 柱浓度出现高低值相差较大的情况，北部和西北部海岸带附近、东南部海域值较高，高值达到 414~416 ppm，而南部近海岸带附近出现 408~410 ppm 的低值。



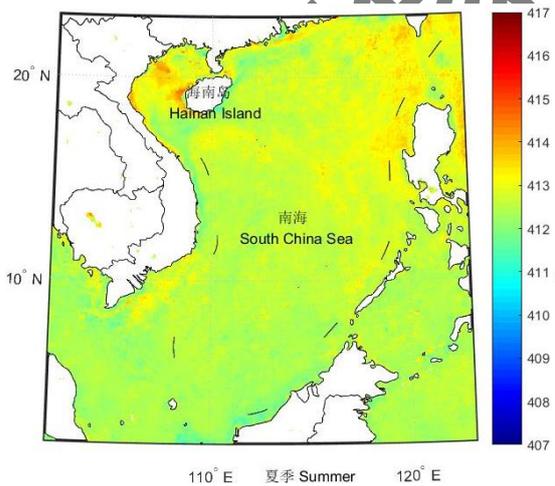
a. OCO-2 数据春季分布 OCO-2 data in spring



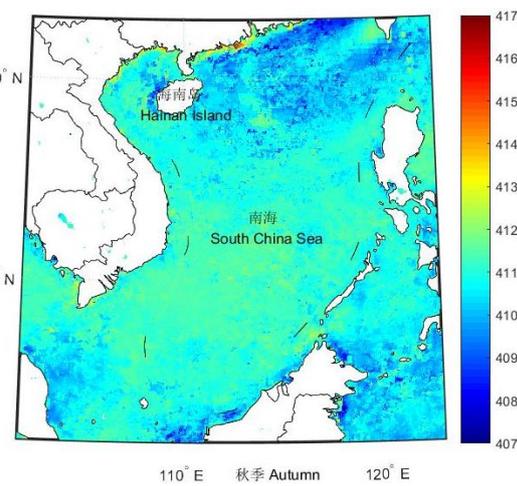
b. 模拟数据冬季分布 Simulated data in winter



c. 模拟数据春季分布 Simulated data in spring



d. 模拟数据夏季分布 Simulated data in summer



e. 模拟数据秋季分布 Simulated data in autumn

图 3 南海大气 CO<sub>2</sub> 柱浓度随季节分布

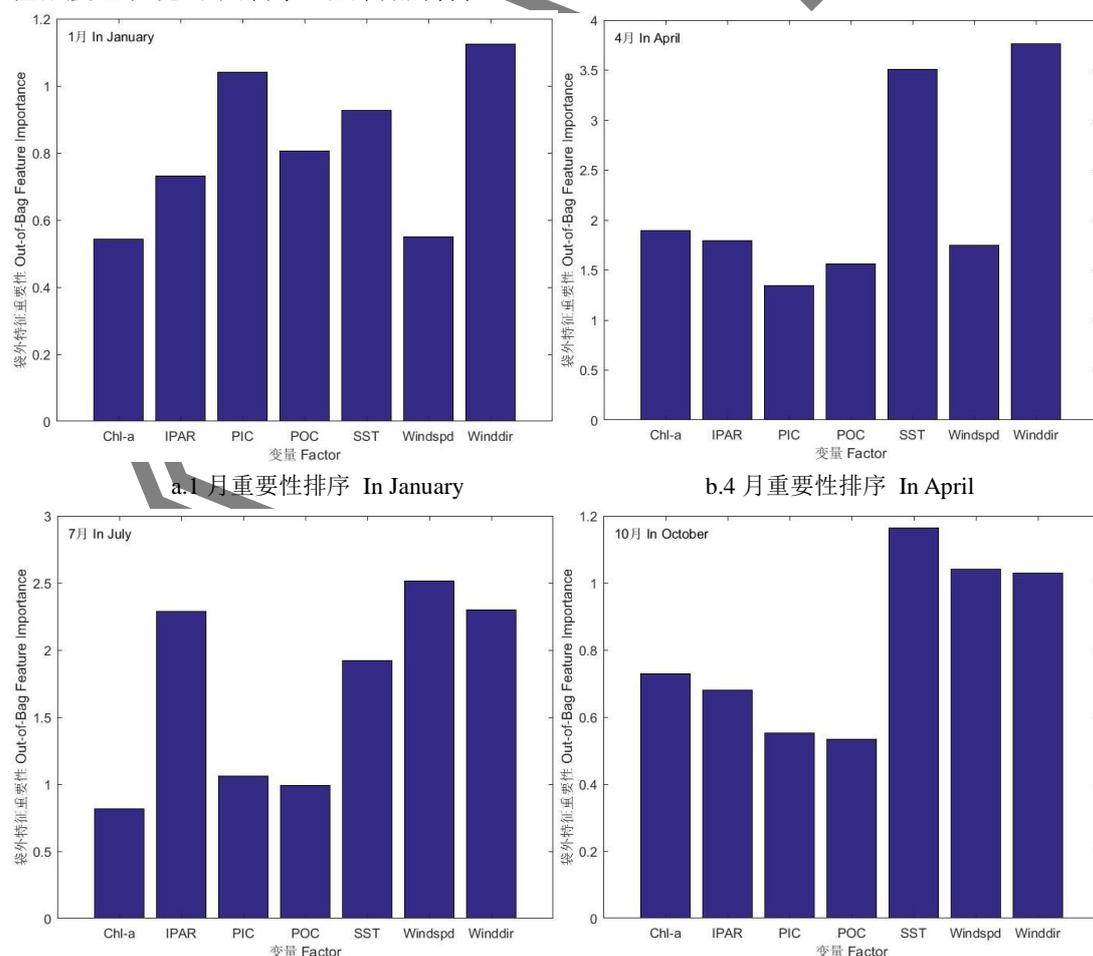
(a. OCO-2 春季, b. 随机森林冬季, c. 随机森林春季, d. 随机森林夏季, e. 随机森林秋季)

Fig. 3 Seasonal distribution of atmospheric CO<sub>2</sub> column concentrations over the South China Sea  
(a. OCO-2 data in spring, b. Simulated data in winter, c. Simulated data in spring, d. Simulated data in summer, e. Simulated data in autumn)

### 3.4 南海大气 CO<sub>2</sub> 柱浓度季节变化原因分析

南海大气 CO<sub>2</sub> 柱浓度季节变化的主要原因有三点：一是“生物泵”随季节变化，即南海真光层的浮游植物的数量、分布和生产效率都随季节发生变化，南海作为陆架边缘海，大陆河流携带的泥沙、污染物、营养盐等进入南海，为浮游植物的繁殖提供了基础，但是不同季节温度的变化、汛期多云对太阳的遮挡等因素又会影响浮游植物的生产效率；二是“溶解度泵”随季节变化，即海表温度本身随季节的变化改变了海—气 CO<sub>2</sub> 分压差，南海会呈现季节性的源—汇转化；三是陆架上空大气 CO<sub>2</sub> 对边缘海的影响，随着季节性风速风向的变化，大陆的高 CO<sub>2</sub> 会被风吹到南海上空，或者原本应由生物泵和溶解度泵影响而产生的 CO<sub>2</sub> 柱浓度差异，在风的影响下产生了空间移位，进而更加复杂的影响海—气 CO<sub>2</sub> 分压。

在不同季节，南海上空 CO<sub>2</sub> 柱浓度受多重因素共同影响，形成明显的季节特征，为了对南海大气 CO<sub>2</sub> 柱浓度季节变化原因进行分析，本文利用随机森林模型内置的 OOB 特征重要性(Out-of-bag feature importance)这个参数来表征每个自变量对随机森林模型精度影响的重要程度，根据 7 个参数的重要性进行排序，结果见图 4。如图所示，1 月和 4 月南海大气 CO<sub>2</sub> 柱浓度的第一影响因素是风向，此时盛行东北季风，来自大陆的高浓度 CO<sub>2</sub> 随东北季风进入南海区域，提高了南海大气 CO<sub>2</sub> 柱浓度，图 3 中展现出的 CO<sub>2</sub> 柱浓度存在明显的北部近海岸 > 中心海盆的这一规律也印证了上述分析；1 月份 PIC 是第二重要影响因素，PIC 多来自陆源输入，所以 1 月份北部近海岸带附近的大气 CO<sub>2</sub> 柱浓度较中心海域明显偏高；7 月份风速和风向是影响最大的 2 个因素，此时南海盛行西南季风，但是北部湾仍以偏北风为主(王春乙, 2014)，且北部湾的风速大于其他海域，叠加因素造成北部湾区域 CO<sub>2</sub> 柱浓度较高；10 月份海温成为最主要影响因素，此时中心海温度高，北部近海岸带温度较低，大气 CO<sub>2</sub> 柱浓度也表现出中部高、北部低的特征。



c.7月重要性排序 In July

d.10月重要性排序 In October

图4 不同月份大气 CO<sub>2</sub> 柱浓度影响因素重要性排序

(a.1月重要性排序, b.4月重要性排序, c.7月重要性排序, d.10月重要性排序)

Fig.4 Histogram of the importance of factors affecting atmospheric CO<sub>2</sub> column concentration in different months  
(a. January, b. April, c. July, d. October)

## 4 结论

4.1 本文基于随机森林模型对多参数、非线性、回归问题的计算优势,提出了一种利用多源卫星的叶绿素 a 浓度、瞬时光和利用率、溶解无机碳、溶解有机碳、海表温度、风速、风向 7 个参数估算南海大气 CO<sub>2</sub> 柱浓度的方法,假设 OCO-2 卫星获得的 XCO<sub>2</sub> 数据为真值,以 2020 年南海数据对模型精度验证, Bias 为 0.27 ppm, R<sup>2</sup> 为 0.59, RMSE 为 1.00 ppm,整体精度令人满意。

4.2 基于宽幅多源遥感数据构建的南海大气 CO<sub>2</sub> 柱浓度估算方法,可以在月或者季的尺度上实现对南海区域的空间全覆盖监测,相较于 OCO-2 卫星的窄幅数据,能够更好的反映南海大气 CO<sub>2</sub> 柱浓度的空间分布特征。

4.3 南海大气 CO<sub>2</sub> 柱浓度呈现明显的季节特征,具体表现为春季>夏季>冬季>秋季。

4.4 在不同的月份,造成南海大气 CO<sub>2</sub> 柱浓度季节差异的主要影响因素是不同的,在 1 月和 4 月,主要影响因素是风向,7 月份风速和风向是影响最大的 2 个因素,10 月份海温成为最主要影响因素。

4.5 考虑到影响南海大气 CO<sub>2</sub> 柱浓度的主要因素随时间变化,下一步拟开展针对不同季节/月份的随机森林南海海上大气 CO<sub>2</sub> 柱浓度估算方法研究。

### 参考文献:

- 白文广,张兴赢,张鹏,2010. 卫星遥感监测中国地区对流层二氧化碳时空变化特征分析[J]. 科学通报, 55(30): 2953-2960. Bai W G, Zhang X Y, Zhang P, 2010. Temporal and distribution of tropospheric CO<sub>2</sub> over China based on satellite observation[J]. Chinese Sci Bull, 55(30): 2953-2960 (in Chinese).
- 蔡大鑫,刘少军,佟金鹤,等,2023. 基于卫星遥感的南海海一气 CO<sub>2</sub> 通量研究[J]. 热带气象学报, 39(4): 462-473. Cai D X, Liu S J, Tong J H, et al, 2023. Study on air-sea CO<sub>2</sub> flux in the South China sea based on satellite remote sensing[J]. J Trop Meteorol, 39(4): 462-473 (in Chinese).
- 邓安健,郭海波,胡洁,等,2020. GOSAT 卫星数据监测中国大陆上空 CO<sub>2</sub> 浓度时空变化特征[J]. 遥感学报, 24(3): 319-325. Deng A J, Guo H B, Hu J, et al, 2020. Analysis of temporal and distribution characteristic of CO<sub>2</sub> concentration over China based on GOSAT satellite data[J]. J Remote Sens, 24(3): 319-325 (in Chinese).
- 葛钰洁,肖钟湧,2022. 基于 GOSAT 卫星的中国 CO<sub>2</sub> 浓度时空特征分析[J]. 中国环境监测, 38(5): 96-108. Ge Y J, Xiao Z Y, 2022. Analysis on the spatial and temporal characteristics of CO<sub>2</sub> concentration in China based on GOSAT satellite[J]. Environ Monitor China, 38(5): 96-108 (in Chinese).
- 杭鑫,曹云,杭蓉蓉,等,2021. 基于随机森林算法与高分观测的太湖叶绿素 a 浓度估算模型[J]. 气象, 47(12): 1525-1536. Hang X, Cao Y, Hang R R, et al, 2021. Estimation model of chlorophyll-a concentration in Taihu Lake based on random forest algorithm and Gaofen observations[J]. Meteor Mon, 47(12): 1525-1536 (in Chinese).
- 加亦瑛,陶明辉,丁思佳,等,2022. 基于卫星遥感的中国地区 XCO<sub>2</sub> 和 XCH<sub>4</sub> 时空分布研究[J]. 大气与环境光学学报, 17(6): 679-692. Jia Y Z, Tao M H, Ding S J, et al, 2022. Spatial and temporal distribution of XCO<sub>2</sub> and XCH<sub>4</sub> in China based on satellite remote sensing[J]. J Atmos Environ Optics, 17(6): 679-692 (in Chinese).
- 李静波,张莹,盖荣丽,2023. 基于机器学习的星载短波红外 CO<sub>2</sub> 柱浓度估算[J]. 中国环境科学, 43(4): 1499-1509. Li J B, Zhang Y, Gai R L, 2023. Estimation of the column concentration of carbon dioxide using spaceborne shortwave infrared spectrometer[J].

- China Environ Sci, 43(4): 1499-1509 (in Chinese).
- 乔年, 卢楚翰, 管兆勇, 等, 2023. CMIP6 模式大气中南北半球水汽质量反相变化: 不同温室气体排放情景差异[J]. 气象学报, 81(5): 776-787. Qiao N, Lu C, Guan Z Y, et al, 2023. Anti-phase relation of water vapor mass between the Northern and Southern Hemispheres in CMIP6 models: differences under different greenhouse gas emission scenarios[J]. Acta Meteorol Sin, 81(5): 776-787 (in Chinese).
- 孙军, 刘东艳, 冯士筵, 2003. 近海生态系统动力学研究中浮游植物采样及分析策略[J]. 海洋与湖沼, 34(2): 224-232. Sun J, Liu D Y, Feng S Z, 2003. Preliminary study on marine phytoplankton sampling and analysis strategy for ecosystem dynamic research in coastal waters[J]. Oceanol Limnol Sin, 34(2): 224-232 (in Chinese).
- 田光辉, 刘少军, 李伟光, 等, 2020. ASCAT 风场在南海的适用性分析[J]. 广东气象, 42(5): 39-43. Tian G H, Liu S J, Li W G, et al, 2020. Analysis of the applicability of ASCAT-derived wind field in the South China Sea[J]. Guangdong Meteor, 42(5): 39-43 (in Chinese).
- 王春乙, 2014. 海南气候[M]. 北京: 气象出版社. Wang C Y, 2014. Hainan Climate[M]. Beijing: China Meteorological Press (in Chinese).
- 王可心, 包云轩, 朱承瑛, 等, 2021. 随机森林回归法在冬季路面温度预报中的应用[J]. 气象, 47(1): 82-93. Wang K X, Bao Y X, Zhu C Y, et al, 2021. Forecasts of road surface temperature in winter based on random forests regression[J]. Meteor Mon, 47(1): 82-93 (in Chinese).
- 叶函函, 王先华, 吴时超, 等, 2021. 高分五号卫星 GMI 大气 CO<sub>2</sub> 反演方法[J]. 大气与环境光学学报, 16(3): 231-238. Ye H H, Wang X H, Wu S C, et al, 2021. Atmospheric CO<sub>2</sub> retrieval method for satellite observations of greenhouse gases monitoring instrument on GF-5[J]. J Atmos Environ Optics, 16(3): 231-238 (in Chinese).
- 詹滨秋, 赵永平, 1989. 二氧化碳, 海洋与气候[J]. 海洋与湖沼, 20(1): 92-99. Zhan B Q, Zhao Y P, 1989. CO<sub>2</sub> oceans and climate[J]. Oceanol Limnol Sin, 20(1): 92-99 (in Chinese).
- 张炳炎, 闫召爱, 郭文杰, 等, 2021. 基于 CO<sub>2</sub> 测量数据的大气辐射传输模型 LBLRTM 优化[J]. 空间科学学报, 41(6): 905-910. Zhang B Y, Yan Z A, Guo W J, et al, 2021. Optimization of atmospheric radiative transfer model LBLRTM based on measured CO<sub>2</sub> data[J]. Chin J Space Sci, 41(6): 905-910 (in Chinese).
- 张远辉, 王伟强, 陈立奇, 2000. 海洋二氧化碳的研究进展[J]. 地球科学进展, 15(5): 559-564. Zhang Y H, Wang W Q, Chen L Q, 2000. Advances in studies of oceanic carbon dioxide[J]. Adv Earth Sci, 15(5): 559-564 (in Chinese).
- Breiman L, 2001. Random Forests[J]. Mach Learn, 45(1): 5-32.
- Crisp D, Pollock H R, Rosenberg R, et al, 2017. The on-orbit performance of the Orbiting Carbon Observatory-2(OCO-2) instrument and its radiometrically calibrated products[J]. Atmos Meas Tech, 10(1): 59-81.
- Guo M, Wang X F, Li J, et al, 2012. Assessment of global carbon dioxide concentration using MODIS and GOSAT data[J]. Sensors, 12(12): 16368-16389, doi: 10.3390/s121216368.
- Guo M, Xu J W, Wang X F, et al, 2015. Estimating CO<sub>2</sub> concentration during the growing season from MODIS and GOSAT in East Asia[J]. Int J Remote Sens, 36(17): 4363-4383, doi: 10.1080/01431161.2015.1081305.
- Hakkarainen J, Jalongo I, Tamminen J, 2016. Direct space-based observations of anthropogenic CO<sub>2</sub> emission areas from OCO-2 [J]. Geophys Res Lett, 43(21): 11400-11406.
- He S C, Yuan Y B, Wang Z H, et al, 2023. Machine learning model-based estimation of XCO<sub>2</sub> with high spatiotemporal resolution in China[J]. Atmosphere, 14(3): 436, doi: 10.3390/atmos14030436.
- He Z H, Zeng Z C, Lei L P, et al, 2017. A data-driven assessment of biosphere-atmosphere interaction impact on seasonal cycle patterns of XCO<sub>2</sub> using GOSAT and MODIS observations[J]. Remote Sens, 9(3): 251.
- Li J, Jia K, Wei X Q, et al, 2022. High-spatiotemporal resolution mapping of spatiotemporally continuous atmospheric CO<sub>2</sub> concentrations over the global continent[J]. Int J Appl Earth Obs Geoinf, 108: 102743.
- Liang A L, Gong W, Han G, et al, 2017a. Comparison of satellite-observed XCO<sub>2</sub> from GOSAT, OCO-2, and ground-based TCCON[J]. Remote Sens, 9(10): 1033, doi: 10.3390/rs9101033.
- Liang A L, Han G, Gong W, et al, 2017b. Comparison of global XCO<sub>2</sub> concentrations from OCO-2 with TCCON data in terms of latitude zones[J]. IEEE J Sel Top Appl Earth Obs Remote Sens, 10(6): 2491-2498.

- Manabe S, Stouffer R J, 1993. Century-scale effects of increased atmospheric CO<sub>2</sub> on the ocean-atmosphere system[J]. *Nature*, 364(6434): 215-218.
- Marland G, Boden T A, Andres R J, 1985. Global, regional, and national fossil fuel CO<sub>2</sub> emissions[Z]. 34-43
- Schneising O, Buchwitz M, Reuter M, et al, 2011. Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY[J]. *Atmos Chem Phys*, 11(6): 2863-2880.
- Tian H Q, Lu C Q, Ciais P, et al, 2016. The terrestrial biosphere as a net source of greenhouse gases to the atmosphere[J]. *Nature*, 531(7593): 225-228.
- Wunch D, Toon G C, Blavier J F L, et al, 2011. The total carbon column observing network[J]. *Phil Trans R Soc A: Math, Phys Eng Sci*, 369(1943): 2087-2112.
- Wunch D, Toon G C, Sherlock V, et al, 2015. The total carbon column observing network's GGG2014 data version[C]//Carbon Dioxide Information Analysis Center. Oak Ridge, Tennessee, USA: Oak Ridge National Laboratory

《气象》待刊