

ZHONG Jia, ZOU Ziming, XU Jiyao, WU Kun, SUN Longchang, YUAN Wei, HU Xiaoyan. Standard Dataset of Ionospheric Equatorial Plasma Bubbles over Southern China Based on Airglow Observations (in Chinese). *Chinese Journal of Space Science*, 2026, **46**(1): 254–264. CSTR:32142.14.cjss.2025–0097. DOI:10.11728/cjss2026.01.2025–0097

## 数据论文

# 基于气辉观测的中国南方电离层赤道 等离子体泡标准数据集

钟佳<sup>1,2</sup> 邹自明<sup>1,2</sup> 徐寄遥<sup>1</sup> 吴坤<sup>3</sup>  
孙龙昌<sup>1</sup> 袁韦<sup>1</sup> 胡晓彦<sup>1,2</sup>

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院国家空间科学数据中心 北京 101407)

3(长沙理工大学物理与电子科学学院 长沙 410114)

**摘要** 研究构建了首个基于气辉观测的电离层等离子体泡标准数据集, 包含等离子体泡事件数据与精确轮廓标注数据. 该数据集源自云南曲靖站 630 nm 波段气辉成像仪历时一个完整太阳活动周 (2012–2022 年) 的连续观测, 所有原始数据均经过图像增强、方位校正、几何畸变校正及地理坐标投影等标准化预处理, 并由专家团队完成等离子体泡事件识别及轮廓标注. 该数据集的时间分辨率高达 3 min, 系统收录了不同太阳活动强度下的等离子体泡事件样本, 涵盖 I 形与 Y 形等多种典型形态. 本数据集为开发高精度监督式人工智能算法提供了高质量基准数据, 有助于促进基于气辉成像的电离层等离子体泡自动化检测与形态演化研究.

**关键词** 电离层赤道等离子体泡, 气辉观测, 标准数据集, 轮廓标注, 中国区域, 太阳活动周, 人工智能

**中图分类号** P352

## 数据库(集) 基本信息

|            |                                                                                                                                                                                                                                   |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据库(集)名称   | 基于气辉观测的中国南方区域电离层等离子体泡标准数据集                                                                                                                                                                                                        |
| 数据通信作者     | 钟佳 (zhongjia@nssc.ac.cn)                                                                                                                                                                                                          |
| 数据作者       | 钟佳, 徐寄遥, 吴坤, 袁韦                                                                                                                                                                                                                   |
| 数据时间范围     | 2012–2022年                                                                                                                                                                                                                        |
| 数据空间范围     | 中国区域电离层; 地理坐标系东经 94°–113°, 北纬 17°–34°                                                                                                                                                                                             |
| 数据分辨率      | 时间分辨率3 min, 空间分辨率2 km                                                                                                                                                                                                             |
| 数据量        | 3.19 GByte, 7022个文件                                                                                                                                                                                                               |
| 数据格式       | *.png, *.json                                                                                                                                                                                                                     |
| 数据库(集)访问网址 | <a href="https://doi.org/10.12176/01.12.015-V01">https://doi.org/10.12176/01.12.015-V01</a>                                                                                                                                       |
| 基金项目       | 国家重点研发计划项目 (2022YFF0711400)                                                                                                                                                                                                       |
| 数据库(集)组成   | 数据集由一系列数据文件组成, 包括等离子体泡图像文件、等离子体泡轮廓标注文件、等离子体泡掩膜图像文件. 等离子体泡图像文件即为人工识别出的包含等离子体泡的图像, 共4664个文件; 轮廓标注文件记录了图像类别 (等离子泡或非等离子体泡)、轮廓像素坐标、像素值, 共2358个文件; 掩膜图像文件将等离子体泡事件区域像素及背景像素进行了二值化处理, 能直观显示等离子体泡, 与轮廓标注文件数量一致. 所有图像大小均为 1024 × 1024 pixel |

2025-06-21 收到原稿, 2025-10-22 收到修定稿

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License  
(<https://creativecommons.org/licenses/by/4.0/>)

# Standard Dataset of Ionospheric Equatorial Plasma Bubbles over Southern China Based on Airglow Observations

ZHONG Jia<sup>1,2</sup>    ZOU Ziming<sup>1,2</sup>    XU Jiyao<sup>1</sup>    WU Kun<sup>3</sup>  
SUN Longchang<sup>1</sup>    YUAN Wei<sup>1</sup>    HU Xiaoyan<sup>1,2</sup>

1(National Space Science Center, Chinese Academy of Sciences, Beijing 100190)

2(National Space Science Data Center, Chinese Academy of Sciences, Beijing 101407)

3(School of Physics and Electronic Science, Changsha University of Science and Technology, Changsha 410114)

**Abstract**    Airglow imaging observations, with their high spatiotemporal resolution and large-scale continuous monitoring capability, provide a crucial approach to studying the fine horizontal structures and evolutionary characteristics of ionospheric equatorial plasma bubbles. However, the current lack of high-quality, professionally annotated plasma bubble datasets severely restricts the application of supervised Artificial Intelligence (AI) algorithms in this field. To address this issue, this study constructs the first standardized dataset of ionospheric plasma bubbles based on airglow observations, including plasma bubble event data products and precise contour annotation data products. The dataset is derived from continuous observations over a full solar activity cycle (2012–2022) by a 630 nm band airglow imager at the Qujing Station in Yunnan, China. All raw data underwent standardized preprocessing, including image enhancement, azimuth correction, geometric distortion correction, and geographic coordinate projection. Expert teams then performed plasma bubble event identification and contour annotation. With a high temporal resolution of 3 minutes, the dataset systematically documents plasma bubble events under varying solar activity intensities, covering multiple typical morphologies such as shaped I- and Y-shaped structures. This dataset provides high-quality benchmark data for developing high-precision supervised AI algorithms, facilitating automated detection and morphological evolution research of ionospheric plasma bubbles based on airglow imaging.

**Key words**    Ionospheric equatorial plasma bubbles, Airglow observation, Standardized dataset, Contour annotation, China region, Solar activity cycle, Artificial Intelligence (AI)

## Database/Dataset Profile

|                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title                        | Standardized Dataset of Ionospheric Equatorial Plasma Bubbles over Southern China Based on Airglow Observations                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Data Communication Author    | ZHONG Jia (zhongjia@nssc.ac.cn)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Data Author (s)              | ZHONG Jia, XU Jiyao, WU Kun, YUAN Wei                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Time Range                   | 2012–2022                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Spatial Range                | Chinese regional ionosphere, 94°E–113°E and 17°N–34°N in the geographical coordinate system                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Data Resolution              | 3 min, 2 km                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Data Volume                  | 3.19 GByte, 7022 files                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Data Format                  | *.png, *.json                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Database/Dataset URL         | <a href="https://doi.org/10.12176/01.12.015-V01">https://doi.org/10.12176/01.12.015-V01</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Foundation Item              | National Key Research and Development Program of China (2022YFF0711400)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Dataset/Database Composition | This dataset comprises a series of data files, including plasma bubble image files, plasma bubble contour annotation files, and plasma bubble mask image files. The plasma bubble image files refer to manually identified images containing plasma bubbles, with a total of 4664 files. The contour annotation files record the image categories (plasma bubble or non-plasma bubble), contour pixel coordinates, and pixel values, totaling 2358 files. The mask images files have binarized the pixels in the plasma bubble event regions and background pixels, which can visually display plasma bubbles; their quantity matches that of the contour annotation files. All images are 1024×1024 pixel in size |

## 0 引言

日落后的低纬度区域电离层由于瑞利-泰勒不稳定性 (Rayleigh-Taylor-like instability), 会形成等离子体密度耗空结构, 即赤道扩展 F 现象或赤道等离子体泡 (Equatorial Plasma Bubble, EPB)<sup>[1-3]</sup>. 研究赤道等离子体泡的形态演化与传播特性, 不仅有助于揭示电离层动力学及热层-电离层耦合机制, 还对保障无线电通信与导航系统的稳定性至关重要.

多种技术手段被用于观测 EPB, 包括卫星原位探测等离子体密度<sup>[4]</sup>、全球导航卫星系统 (Global Navigation Satellite System, GNSS) 反演的总电子含量<sup>[5]</sup>、雷达回波观测<sup>[6]</sup>和数字测高仪频高图<sup>[7]</sup>等. 其中地基全天空空气辉成像仪气辉成像观测凭借其高时空分辨率与大范围连续监测能力, 为研究电离层等离子体泡的水平精细结构及其演化特征提供了重要手段. 然而, 该技术依赖于光学观测, 易受到云层覆盖、月光干扰等天气条件的限制, 可能影响观测数据中等离子体泡图像连续性与可见性. 国内外利用全天空空气辉成像仪, 开展了等离子体泡发生率、形态特征及传播特征<sup>[8-13]</sup>等系列研究工作. 然而, 现有数据分析方法多依赖于人工识别及短期数据分析, 且主要集中于北美、巴西和印度等地区. 这种研究现状导致对中国扇区 EPB 演化规律认识不足, 特别是缺乏太阳活动周时间尺度的特征分析. 究其原因, 主要受限于传统数据处理方法的效率瓶颈. 以中国子午工程<sup>[14]</sup>为例, 其单台站单通道气辉成像仪年数据输出量已达约数百 GB. 随着子午工程二期的建成, 数十个观测设备投入运行<sup>[15]</sup>, 人工处理海量观测数据面临巨大困难.

人工智能技术处理海量图像数据具有天然技术优势, 能极大提升数据分析效率<sup>[16-18]</sup>. 特别是在气辉图像数据分析方面已取得了部分进展, 例如对重力波事件<sup>[19]</sup>、行进式电离层扰动事件<sup>[20]</sup>分析. 标准数据集是数据科学研究的基石, 其为人工智能算法开发提供统一的评估基准, 确保研究结果的可比性与可复现性, 同时显著降低数据采集与清洗成本, 加速模型迭代. 然而, 当前缺乏高质量、专业标注的等离子体泡数据集, 这严重制约了监督式人工智能算法在等离子体泡研究中的应用.

为此, 研究构建了首个基于气辉观测的中国南方区域电离层等离子体泡标准数据集, 包含等离子体泡

事件数据和精确轮廓标注数据. 该数据集源自云南曲靖站 630 nm 波段气辉成像仪历时一个完整太阳活动周 (2012—2022 年) 的连续观测, 所有原始数据均经过图像增强、方位校正、几何畸变校正及地理坐标投影等标准化预处理, 并由专家团队完成 2012—2022 年等离子体泡事件识别及 2012—2013 年的等离子体泡轮廓标注, 系统收录了不同太阳活动强度下的等离子体泡事件样本, 涵盖 I 形和 Y 形等多种典型形态. 通过规范化的标注及质量控制, 等离子体泡标准数据集为开发电离层领域高精度监督式人工智能算法提供了高质量基准数据, 有助于促进基于气辉成像的电离层等离子体泡自动化检测与形态演化研究.

## 1 数据采集与处理方法

### 1.1 数据采集方法

原始数据采集自云南曲靖站 (地理坐标: 25°N, 104°E; 地磁坐标: 15.1°N, 176.7°E) 双通道全天空空气辉成像仪 OI 气辉辐射强度的二维强度分布图像 (每个 CCD 像元记录的光子数), 图 1 给出了观测台站位置及观测视野覆盖的有效空间范围.

全天空空气辉成像仪由中国科学院国家空间科学中心自主研发. 图 2 和 3 给出了全天空空气辉成像仪内

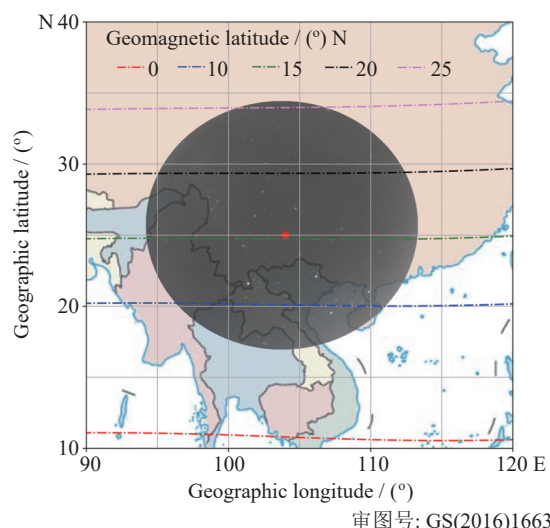


图 1 曲靖站 (红色点) 全天空空气辉成像仪 160°有效观测视场角覆盖空间范围

Fig. 1 Spatial coverage of the 160° effective Field-of-View (FOV) for the all-sky airglow imager at Qujing Station (Red dot)

部组成结构, 主要包括鱼眼镜头、透镜系统、滤波片组和电热制冷 CCD 相机. 鱼眼镜头的视角为 $180^\circ$ , 滤波片采用 OI 630 nm 波段的滤波片, CCD 相机参数为  $1024 \times 1024$  pixel, 像元面积为  $13 \mu\text{m} \times 13 \mu\text{m}$ , 暗电流小于  $0.05 \text{ pixel}^{-1} \cdot \text{s}^{-1}$ . 光路图显示全天空气辉成像仪的视场角较大, 但是焦距较短, 会造成观测图像具有明显的视场畸变.

全天空气辉成像图曝光时间 180 s, 实际观测波长 630 nm, 对应峰值高度约为  $(250 \pm 50) \text{ km}$ , 图像尺寸为  $1024 \times 1024$  pixel, 像素深度为 16 位, 如图 4 所示.

### 1.2 数据处理方法

针对原始气辉观测图像, 经过图 5 所示的加工处理流程, 生成等离子体泡标准数据集.

数据集具体加工处理流程如下.

(1) 图像增强与方位校正. 应用直方图均衡化算法增强等离子体泡特征. 夜气辉成像属于弱光探测范畴, 其信号强度极低, 在通过 2 nm 窄带滤光片后, 原始图像的灰度值普遍集中于较低区间, 直接观察时, 图像整体偏暗, 难以辨识等离子体泡结构. 因此, 需通过适当的亮度拉伸处理 (即图像均衡化) 来增强图像对比度, 从而使等离子体泡信息得以显现. 图像对比

度可由灰度级动态范围表征, 借助算法调整各像元的灰度值, 可实现对比度的有效提升. 主要处理流程包括: 首先计算图像的灰度直方图; 进而得到该直方图的累积分布; 确定累积分布中像素占比低于 0.01 与高于 0.99 的临界灰度值, 将低于 0.01 的部分映射至 0, 高于 0.99 的部分映射至 255; 在此基础上, 建立输入与输出灰度级之间的映射函数, 并依据该函数逐像素转换灰度值. 最终, 将 16 位图像线性压缩至 8 位动态范围内. 16 位深度图像线性压缩至 8 位后, 不影响后续等离子体泡图像筛选, 且能降低存储. 最后旋转图像, 使地理南北极走向与直角坐标系吻合, 正北向上.

(2) 候选等离子体泡图像筛选. 基于图像亮度梯度变化以及等离子体泡发生规律的时间段, 人工筛选

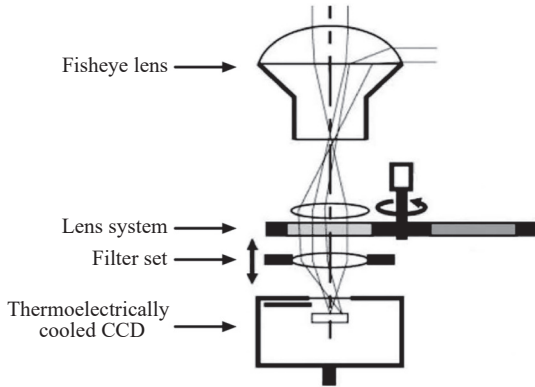


图 3 全天空气辉成像仪内部结构<sup>[21]</sup>

Fig. 3 All-sky airglow imager internal structure



图 2 全天空气辉成像仪实物<sup>[21]</sup>

Fig. 2 Picture of the all-sky airglow imager

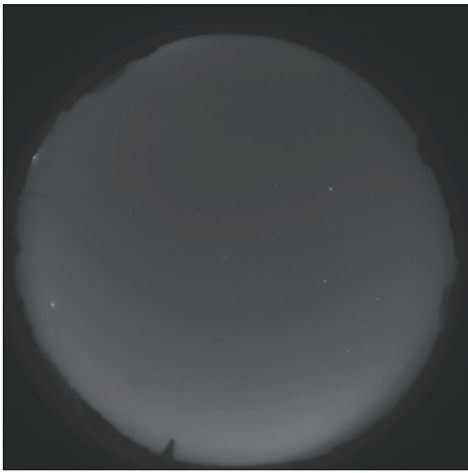


图 4 全天空气辉成像仪原始观测数据

Fig. 4 Raw observational data of the all-sky airglow imager

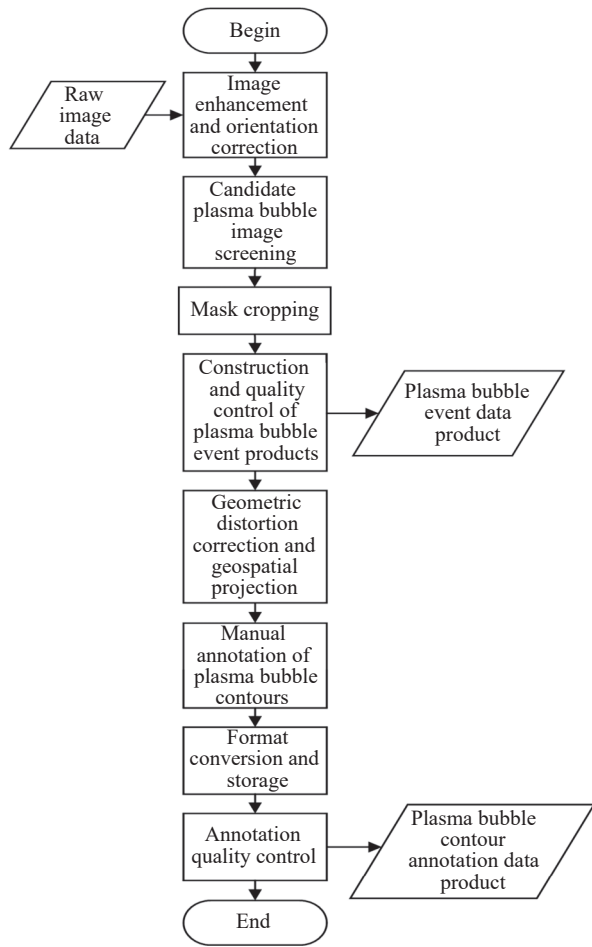


图 5 数据集加工流程

Fig. 5 Dataset processing pipeline

数据中包含等离子体泡的图像,并覆盖 I 形与 Y 形等多种形态等离子体泡,保证数据集样本的丰富性。

(3) 掩膜裁剪. 由于鱼镜头成像存在畸变,气辉观测图像在天顶区域的空间分辨率最高,随着视场角远离天顶,分辨率逐渐降低,单位像素对应的实际径向距离也随之增大. 考虑到图像边缘区域存在显著的分辨率下降及几何形变,通常将有效观测视场设定为  $160^\circ$ ,并假设气辉层呈圆弧状分布. 为消除无效区域的影响,对图像进行圆形掩膜处理:以图像中心为原点,设置掩膜尺寸为  $400 \times 400$  pixel,保留圆形区域内的像素值不变,将外部区域像素统一置零,从而实现有效视场的提取与畸变区域的剔除。

(4) 等离子泡事件数据构建与质量控制. 基于裁剪后的图像,经过领域专家再次确认等离子体泡的存在,为保证样本质量及其代表性,从中筛选出最终的图像作为等离子体泡事件数据。

(5) 几何畸形校正与地理空间投影. 为了消除鱼镜头畸变影响,并将气辉图像投影至均匀地理网格<sup>[22]</sup>,方便后续求解与分析其物理参量. 将图像中的像素坐标  $(i, j)$  和星图中的标准坐标  $(f, g)$  用如下关系表示:

$$\begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} 1 \\ i \\ j \end{bmatrix}. \quad (1)$$

$(f, g)$  可以与方位角  $u$  和高度角  $e$  建立如下联系

$$\begin{bmatrix} f \\ g \end{bmatrix} = G(e) \times \begin{bmatrix} \sin u \\ \cos u \end{bmatrix}. \quad (2)$$

其中,  $a_0, a_1, a_2, b_0, b_1, b_2$  分别为待拟合的参数;  $G(e)$  为透镜函数,表示归一化的天顶距  $G$  与高度角  $e$  的函数,由全天空气辉成像仪透镜决定,通常情况由厂商给出. 此处考虑  $G(e)$  为  $e$  的一阶多项式拟合结果为

$$G(e) = 1 - e/90. \quad (3)$$

高度-方位坐标  $(e, u)$  通过几何转换即可以转变为实际的地理坐标  $(x, y)$ 。

通过星图软件 Stellarium 人工识别,并对比星图中恒星坐标与气辉图像中的像素坐标,从中选取 100 个星点作为校准与地理投影的基准. 在地理坐标投影过程中,设定每个像素代表的空间距离为 2 km. 具体过程如下。

(1) 人工标注等离子体泡轮廓. 利用 labelme 软件人工逐像素点标注图像中等离子体泡轮廓,沿等离子体泡边缘精确绘制多边形,包含等离子体泡核心及边缘过渡区域. 将多边形包含的区域类别设置为等离子体泡,剩下图像区域设置为背景。

(2) 格式转换与存储. 将标注后的等离子体泡图像存储为对应的 PNG 格式的图像文件和 JSON 格式的标注文件。

(3) 标注质量控制. 针对等离子体泡轮廓结果图,经过专家逐张检查,查看标注的轮廓是否精细、准确,确保明显的等离子体泡结构均被标注,去除误标注的图像,确定前后标注的一致性,最后保留高质量标注结果图像,存为最终的等离子体泡轮廓标注数据。

由于等离子体泡事件数据仅需从图像中辨别是否有等离子体泡,不需要提取物理特征参量,所以其数据处理步骤不包含几何畸形校正与地理空间投影. 从数据集加工流程图中,选择前四个处理步骤即可. 处理步骤中,图像增强与方位校正后的样本可初步查

看是否有等离子体泡, 经过掩膜裁剪, 去除边缘大量畸变及环境干扰信息, 保留图像核心区域信息. 经过初筛和二次确认, 保证了等离子体泡事件数据的质量可靠. 最终, 获得了高质量事件图像样本.

等离子体泡轮廓标注数据主要用于支持获取等离子体泡物理特征参量, 相比于等离子体泡事件数据, 需要增加几何畸形校正与地理空间投影、人工标注等离子体泡轮廓等处理步骤. 通过几何畸形校正与地理空间投影, 将原本因鱼镜头畸变效应影响的图像进行校正, 并投影至地理空间, 使所有像素代表的实际空间距离相等, 从而获得准确的形态特征. 经过人工精细化标注等离子体泡区域轮廓, 获取事件区域的像素坐标信息, 生成了高质量的标注文件及等离子体泡掩膜文件, 其中每张等离子体泡图像对应一个 json 格式标注文件及一张等离子体泡掩膜图像.

本数据集构建过程严格遵循科学数据处理规范, 所有处理步骤均可通过提供的技术细节复现. 数据集可用于等离子体泡形态学分析、机器学习模型训练等研究领域.

## 2 数据样本描述

### 2.1 数据记录

本数据集包括等离子体泡事件数据与等离子体泡轮廓标注数据, 图像尺寸均为  $1024 \times 1024$  pixel, 时间分辨率 3 min. 等离子体泡事件数据时间范围为 2012 年 1 月至 2022 年 12 月, 文件个数为 4664, 数据量 2.81 GByte; 等离子体泡轮廓标注数据时间范围为 2012 年 1 月至 2013 年 3 月, 文件个数 2358, 数据量 390 MByte.

图 6 给出了 2014 年 3 月 1 日夜晩观测到的等离子体泡事件. 该图像方位为上北下南左西右东, 图中所示的观测视野中灰度相对于背景区域明显偏暗的地方为等离子体泡事件区域, 其中包含多个等离子体泡, 且等离子体泡呈现复杂的分叉结构.

等离子体泡轮廓标注数据是经过预处理后形成的待标注等离子体泡观测图像、人工标注的 JSON 格式标注文件及等离子体泡掩膜图像三类文件, 每类文件数量均为 786 个, 数据样本覆盖简单 I 形、复杂 Y 形等多种形态的等离子体轮廓标注, 数据样本具体如图 7~9 所示.

图 7~9(a) 给出了待标注的等离子体泡图像, 该

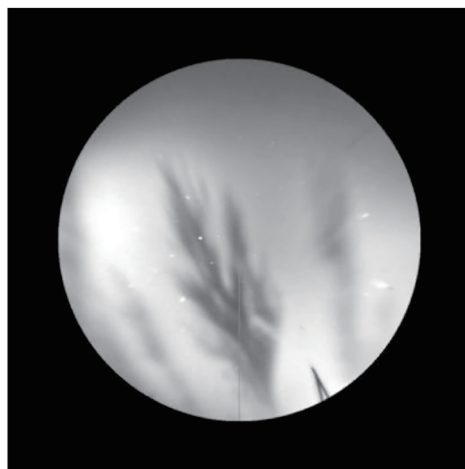


图 6 曲靖站 2014 年 3 月 1 日 23:42:01 (UTC+8) 的等离子体泡事件图像

Fig. 6 Plasma bubble event image observed at Qujing Station at 23:42:01 (UTC+8) on 1 March 2014

图像中每个像素代表实际空间距离为 2 km; 图 7~9(b) 为等离子体泡掩膜图像, 该图像基于多边形标注生成, 其中红色区域代表等离子体泡目标区域, 黑色区域表示背景. 图 10 为等离子体泡轮廓标注文件, 文件内容包括红色多边形包含的标注区域类别 plas-mabubble, 标注区域轮廓像素点横纵坐标, 例如 [574.88, 675.66].

### 2.2 数据集目录结构

等离子体泡标准数据集包括等离子体泡事件数据与等离子体泡轮廓标注数据. 等离子体泡事件数据由等离子体泡图像文件与非等离子体泡图像文件组成. 等离子体泡轮廓标注数据文件夹由 images, annotations, masks 三个主要文件夹组成. 三个文件夹下分别存储待标注的等离子体泡图像、等离子体泡标注文件及等离子体泡掩膜图像文件, 所有图像均以 PNG 格式提供, 标注文件以 json 格式提供. 等离子体泡标准数据集文件组织形式如图 11 所示.

## 3 数据质量控制与评估

本数据集在标注过程中, 首先由一名经过专业培训的标注员对所有图像进行统一标注, 标注结果再经过一名电离层领域专家检查及确认, 剔除标注质量较低的数据样本, 最后由一名电离层领域专家进行抽查, 从事件准确性评估、样本完整性、标注样本的轮廓准确性、样本形态丰富性方面分别进行评估, 确认

最终的等离子体泡标准数据集的高质量与可靠性. 具体过程如下.

(1) 数据质量控制. 数据采集过程中时间连续性较好, 整体保持 3 min 的时间分辨率, 仅在极少数日期存在缺失情况. 此外, 针对原始图像中大量受云、月光污染或过曝光的低质量图像, 可以通过人工筛除此类质量较低的图像, 避免此类图像干扰后续等离子体泡的识别任务.

(2) 事件准确性评估. 采用专业培训人员初筛和领域专家二次审核, 并根据等离子体泡事件发生规律及气辉图像灰度梯度变化, 准确筛选出等离子体泡事件样本.

(3) 样本完整性评估. 等离子体泡发生率受到太

阳活动水平的影响, 为了获得不同太阳活动水平下等离子体泡样本. 选择的等离子体泡事件图像覆盖一个完整太阳活动周 (2012—2022 年). 表 1 给出了不同年份的观测夜晚数量、无光学污染的观测夜晚数量、气辉观测图像数量、质量较好的干净气辉图像数量及等离子体泡样本数量的分布情况. EPB 样本逐年逐月分布与太阳活动水平分布呈现良好的正相关性, 如图 12 和图 13 所示 (图 12 中橙色曲线代表等离子体泡样本数量, 黄色竖条代表太阳活动水平; 图 13 中绿色数据点上方数字表示特定年份, 蓝色直线表示数据点线性拟合结果). 这与以往的研究结论一致<sup>[23,24]</sup>, 说明本数据集中等离子体泡样本分布能够准确反映等离子体泡事件的实际分布特征. 因此, 数据集中的样

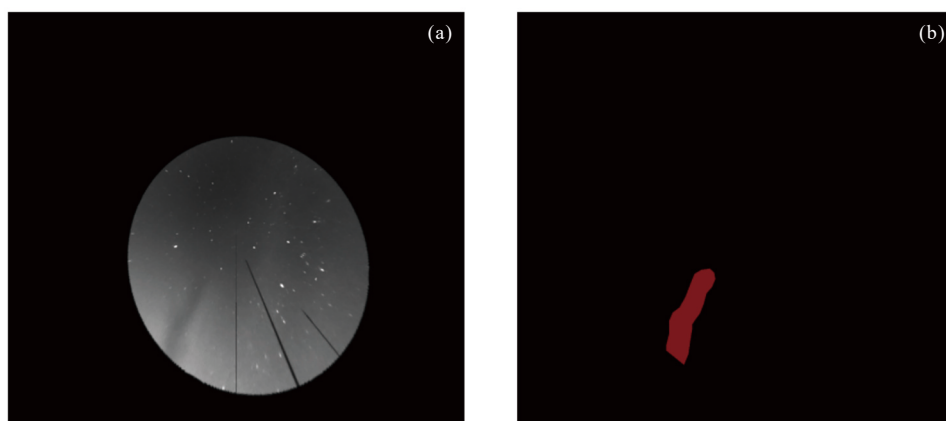


图 7 简单 I 形等离子体泡事件的轮廓标注. (a) 待标注的等离子体泡观测图像, (b) 等离子体泡掩膜图像

Fig. 7 Contour annotation of a simple I-shaped plasma bubble event. (a) Observation image of plasma bubble to be annotated, (b) mask image of plasma bubble

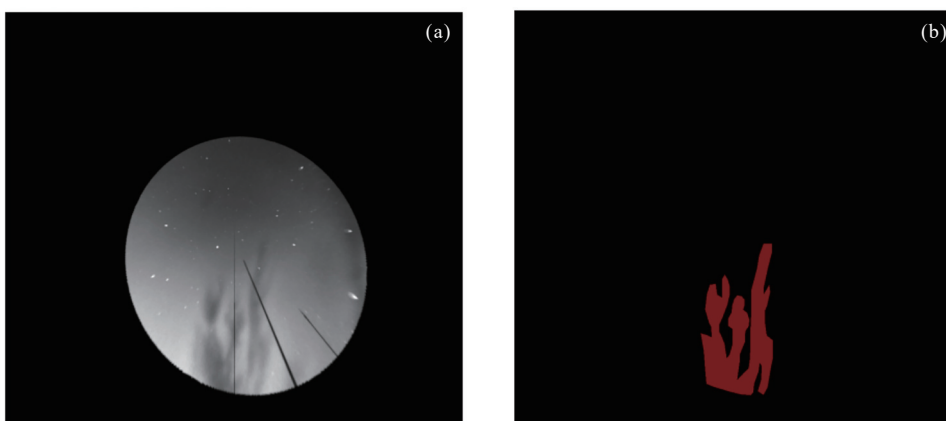


图 8 具有复杂分叉结构的 Y 形等离子体泡事件轮廓标注. (a) 待标注的等离子体泡观测图像, (b) 等离子体泡掩膜图像

Fig. 8 Contour annotation of a Y-shaped plasma bubble event with complex bifurcation structure.

(a) Observation image of plasma bubble to be annotated, (b) mask image of plasma bubble

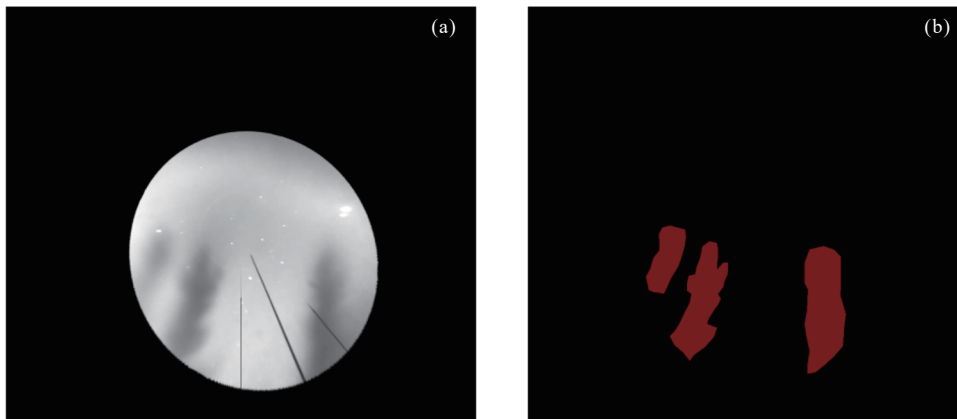


图 9 多个 I 形与 Y 形离子体泡事件轮廓标注. (a) 待标注的等离子体泡观测图像, (b) 等离子体泡掩膜图像  
Fig. 9 Contour annotation of multiple I-shaped and Y-shaped plasma bubble events. (a) Observation image of plasma bubble to be annotated, (b) mask image of plasma bubble

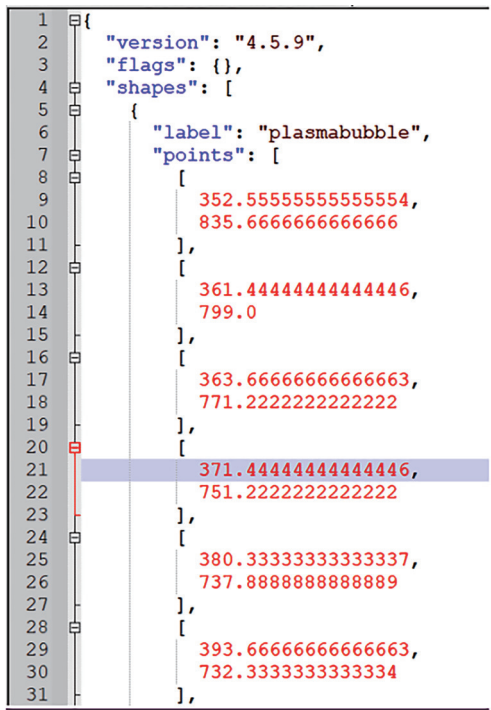


图 10 等离子体泡事件轮廓标注文件  
Fig. 10 Contour annotation file of plasma bubble events

本具有较好的完整性,可有效反映不同太阳活动水平下的等离子体泡事件.

为了进一步验证数据集的质量,从样本丰富性与轮廓准确性两个维度开展评估.

(1) 样本丰富性评估. 标注样本中需包含简单 I 形、复杂分叉结构 Y 形,以及多种混合形态结果的

等离子体泡样本.

(2) 轮廓准确性评估. 基于领域专家经验判断,标注的轮廓是否符合完整的等离子体泡核心区域及边缘过渡区域. 最终确认了所有的等离子体泡轮廓标注结果具有高精度与可靠性.

## 4 数据价值

目前,国内外暂无公开的气辉图像等离子体泡标准数据集. 本数据集具有数据来源与覆盖范围的独特性及潜在应用领域的扩展价值.

(1) 数据来源与覆盖范围的独特性. 与巴西等低纬度台站相比,本数据集来自中国中低纬度关键区(地磁 15.1°N)的自主观测,填补了东亚扇区系统性 EPB 标注数据的空白. 相较于美国 C/NOFS 卫星的穿越式观测或 DMSP/SSUSI 的全球扫描数据,本数据集具有连续 3 min 高时间分辨率(同类影像数据集通常为 15~30 min),能捕捉等离子体泡的快速演化过程.

(2) 潜在应用领域的扩展价值. 可为机器学习提供基准数据集,包含 4664 张等离子体泡图像和 786 张等离子体泡轮廓标注图像,明确区分等离子体泡本体与背景. 支持多尺度研究,包括单个等离子体泡(约 10 km)及等离子体泡群(约 1000 km).

本数据集准确反映了  $(250\pm50)$  km 高度的中高层大气气辉辐射强度分布,在水平方向上观测等离子体泡形态变化,具备较高的时空分辨率. 数据集包含了不同形态、不同空间尺度的等离子体泡图像及其标

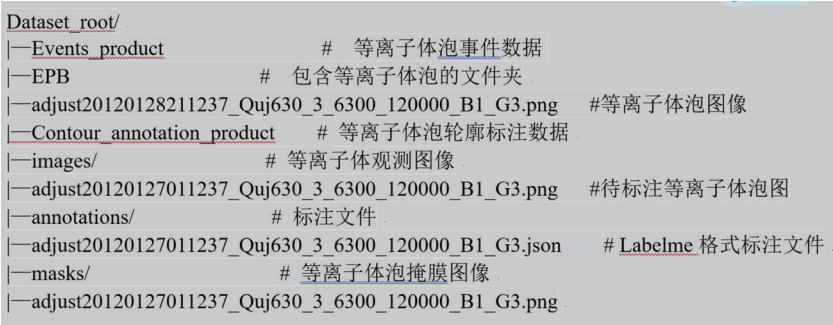


图 11 等离子体泡标准数据集文件组织形式  
Fig. 11 File organization format of plasma bubble standard dataset

表 1 曲靖台站不同年份观测夜晚与气辉观测图像及等离子体泡样本分布  
Table 1 Distribution of observation nights, airglow observation images and plasma bubble samples at Qujing Station in different years

| Year  | Observation nights | Clear nights | Images | Good quality images | EPB images |
|-------|--------------------|--------------|--------|---------------------|------------|
| 2012  | 291                | 108          | 43389  | 13639               | 474        |
| 2013  | 347                | 132          | 58441  | 16012               | 799        |
| 2014  | 357                | 145          | 68662  | 18066               | 1893       |
| 2015  | 360                | 121          | 75368  | 19112               | 889        |
| 2016  | 323                | 112          | 65849  | 18012               | 99         |
| 2017  | 341                | 110          | 71820  | 17739               | 0          |
| 2018  | 313                | 129          | 64227  | 19510               | 21         |
| 2019  | 294                | 133          | 60604  | 18701               | 0          |
| 2020  | 313                | 134          | 62052  | 17150               | 25         |
| 2021  | 363                | 135          | 77322  | 20012               | 20         |
| 2022  | 355                | 133          | 75641  | 20124               | 444        |
| Total | 3657               | 1392         | 723375 | 198077              | 4664       |

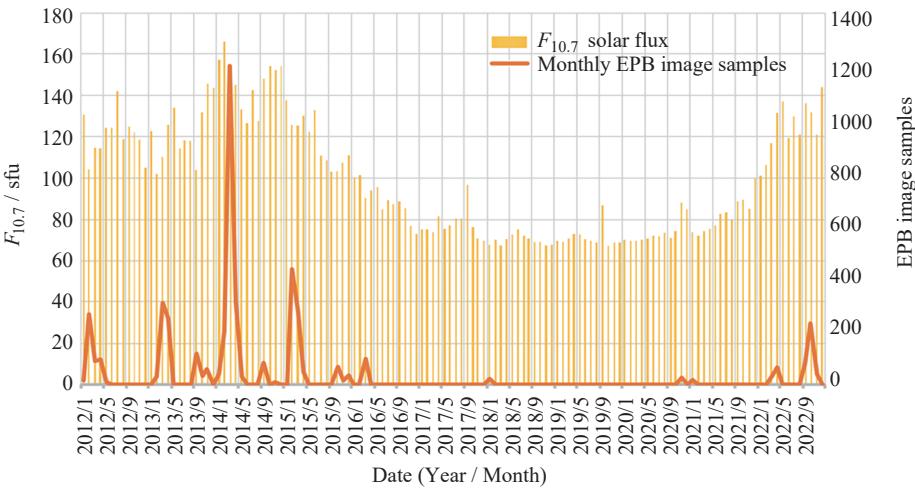


图 12 等离子体泡样本逐年逐月分布与太阳活动水平的关系<sup>[25]</sup>  
Fig. 12 Relationship between interannual and monthly distributions of plasma bubble samples and solar activity levels



图 13 等离子体泡图像数量年均值与太阳射电流量年均值的关系

Fig. 13 Relationship between the annual mean of EPB number image and the annual mean of solar flux

注结果, 为研究等离子体泡的触发机制、东亚地区空间天气效应、以及低纬电离层建模提供了不可替代的基础数据支撑, 特别在机器学习应用与中小尺度结构分析方面具有独特优势。

## 5 数据使用方法与建议

通过提供经过整理的 2012—2022 年历时一个完整太阳活动周的等离子体泡气辉图像事件与 2012—2013 年等离子体泡轮廓标注数据, 旨在为大规模气辉图像中电离层扰动事件研究提供便利。目前, 从未发布过如此长时间跨度的事件图像。通过这一补充, 期望计算机视觉社区和电离层物理领域能利用该等离子体泡标准数据集开展相关科学和应用研究。

### 5.1 大规模气辉图像中等离子体泡事件识别

本数据集首次提供了中国区域覆盖完整太阳活动周的等离子体泡事件样本, 可为电离层物理领域的研究提供重要支撑——此前该领域尚未有如此全面的公开样本资源可供使用。此外, 为了构建一个高效、可扩展的气辉图像等离子体泡事件识别系统<sup>[25]</sup>, 以支持研究人员在大规模地基全天空气辉图像存储库 (目前包含超过千万张图像) 中检索目标样本, 需依赖大规模带标签的图像数据开展模型训练, 而本数据集适配计算机视觉任务应用, 可直接用于等离子体泡事件机器识别模型的训练与验证。

### 5.2 气辉图像中感兴趣区域的发现

作为计算机视觉与图像检索领域的另一个热门

研究方向, 感兴趣区域的检测是一类高度依赖大规模带标签训练语料库支撑的典型任务。在前期研究<sup>[25]</sup>中, 探索了如何利用有监督学习技术, 在气辉图像里定位包含潜在等离子体泡的感兴趣区域。通过对多种方法进行实验, 最终构建出一套可自动检测等离子体泡事件、并提取其面积、纬向扩展、漂移速度等物理特征的有效方案。这类研究工作的开展, 离不开带标签的数据集的支撑。

### 5.3 气辉图像与其他观测设备数据的比较

使用卫星原位探测、GPS-TEC、雷达观测、数字测高仪等电离层观测设备均可以观测等离子体泡事件, 但是不同设备探测原理以及观测设备的位置差异, 均只能观测到等离子体泡事件剖面特征或局部结构。因此, 开展多设备的联合观测分析, 可以获得更准确、更全面的等离子体泡事件演化信息, 例如结合 GPS-TEC 数据与气辉成像观测数据分析等离子体泡的传播特性<sup>[21,26]</sup>。所提供的基于气辉观测的等离子体泡数据集, 可促进该类研究工作的开展。

## 6 数据可用性声明

数据已汇交至中国科学院国家空间科学中心进行公开共享服务, 共享协议 (访问权限) 为 CCBY 4.0。共享途径为线下共享途径, 共享范围为有条件共享, 用户可在线提交使用申请, 经审核通过后方可获取数据。

**致谢** 研究使用的气辉数据由国家重大科技基础设施子午工程提供, 云南曲靖站给予了大力赞助。

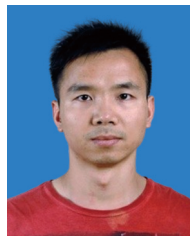
**利益冲突** 所有作者均声明不存在利益冲突。

## 参考文献

- [1] OTT E. Theory of Rayleigh-Taylor bubbles in the equatorial ionosphere[J]. *Journal of Geophysical Research: Space Physics*, 1978, **83**(A5): 2066-2070
- [2] WEBER E J, BUCHAU J, EATHER R H, *et al.* North-south aligned equatorial airglow depletions[J]. *Journal of Geophysical Research: Space Physics*, 1978, **83**(A2): 712
- [3] KELLEY M C. The Earth's Ionosphere: Plasma Physics and Electrodynamics, Second Edition[M]. Boston: Academic Press, c2009. Hardback ISBN: 9780120884254
- [4] AGGSON T L, LAAKSO H, MAYNARD N C, *et al.* In situ observations of bifurcation of equatorial ionospheric plasma depletions[J]. *Journal of Geophysical Research: Space Physics*, 1996, **101**(A3): 5125-5132

- [5] IMOÇIN E, INYURT S, TEMUÇIN H, *et al.* Investigation of equatorial plasma bubble irregularities under different geomagnetic conditions during the equinoxes and the occurrence of plasma bubble suppression[J]. *Acta Astronautica*, 2020, **177**: 341-350
- [6] WOODMAN R F, LA HOZ C. Radar observations of F region equatorial irregularities[J]. *Journal of Geophysical Research*, 1976, **81**(31): 5447-5466
- [7] WOODMAN R F. Spread F-an old equatorial aeronomy problem finally resolved[J]. *Annales Geophysicae*, 2009, **27**(5): 1915-1934
- [8] MENDILLO M, BAUMGARDNER J. Airglow characteristics of equatorial plasma depletions[J]. *Journal of Geophysical Research: Space Physics*, 1982, **87**(A9): 7641-7652
- [9] SAHAI Y, FAGUNDES P R, BITTENCOURT J A. Transequatorial F-region ionospheric plasma bubbles: solar cycle effects[J]. *Journal of Atmospheric and Solar-Terrestrial Physics*, 2000, **62**(15): 1377-1383
- [10] KELLEY M C, MAKELA J J, LEDVINA B M, *et al.* Observations of equatorial spread-F from Haleakala, Hawaii [J]. *Geophysical Research Letters*, 2002, **29**(20): 2003
- [11] MAKELA J J, LEDVINA B M, KELLEY M C, *et al.* Analysis of the seasonal variations of equatorial plasma bubble occurrence observed from Haleakala, Hawaii[J]. *Annales Geophysicae*, 2004, **22**(9): 3109-3121
- [12] TAORI A, SINDHYA A. Measurements of equatorial plasma depletion velocity using 630 nm airglow imaging over a low-latitude Indian station[J]. *Journal of Geophysical Research: Space Physics*, 2014, **119**(1): 396-401
- [13] NARAYANAN V L, GURUBARAN S, SHINY M B B, *et al.* Some new insights of the characteristics of equatorial plasma bubbles obtained from Indian region[J]. *Journal of Atmospheric and Solar-Terrestrial Physics*, 2017, **156**: 80
- [14] WANG C. New chains of space weather monitoring stations in China[J]. *Space Weather*, 2010, **8**(8): S08001
- [15] LI Q Z, XU J Y, GUSMAN A R, *et al.* Upper-atmosphere responses to the 2022 Hunga Tonga-Hunga Ha'apai volcanic eruption *via* acoustic gravity waves and air-sea interaction[J]. *Atmospheric Chemistry and Physics*, 2024, **24**(14): 8343-8361
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, **60**(6): 84-90
- [17] SMIRNOV E A, TIMOSHENKO D M, ANDRIANOV S N. Comparison of regularization methods for ImageNet classification with deep convolutional neural networks[J]. *AASRI Procedia*, 2014, **6**: 89-94
- [18] LIU L, OUYANG W L, WANG X G, *et al.* Deep learning for generic object detection: a survey[J]. *International Journal of Computer Vision*, 2020, **128**(2): 261-318
- [19] LAI C, XU J Y, YUE J, *et al.* Automatic extraction of gravity waves from all-sky airglow image based on machine learning[J]. *Remote Sensing*, 2019, **11**(13): 1516
- [20] LAI C, XU J Y, LIN Z S, *et al.* Statistical characteristics of nighttime medium-scale traveling ionospheric disturbances from 10-years of airglow observation by the machine learning method[J]. *Space Weather*, 2023, **21**(5): e2023SW-003430
- [21] SUN Longchang. Study on Low- and Mid-Latitudinal Ionospheric Irregularity Based on Ground-Based and Satellite Observations[D]. Beijing: University of Chinese Academy of Sciences (National Space Science Center, Chinese Academy of Sciences), 2017: 47-48
- [22] GARCIA F J, TAYLOR M J, KELLEY M C. Two-dimensional spectral analysis of mesospheric airglow image data[J]. *Applied Optics*, 1997, **36**(29): 7374-7385
- [23] HUANG C Y, BURKE W J, MACHUZAK J S, *et al.* Equatorial plasma bubbles observed by DMSP satellites during a full solar cycle: toward a global climatology[J]. *Journal of Geophysical Research: Space Physics*, 2002, **107**(A12): 1434
- [24] WU K, XU J Y, ZHU Y J, *et al.* Occurrence characteristics of branching structures in equatorial plasma bubbles: a statistical study based on all-sky imagers in China[J]. *Earth and Planetary Physics*, 2021, **5**(5): 407-415
- [25] ZHONG J, ZOU Z M, WU K, *et al.* Automatic detection and feature extraction of equatorial plasma bubbles from all-sky airglow image based on machine learning[J]. *Space Weather*, 2025, **23**(5): e2025SW004336
- [26] HAASE J S, DAUTERMANN T, TAYLOR M J, *et al.* Propagation of plasma bubbles observed in Brazil from GPS and airglow data[J]. *Advances in Space Research*, 2011, **47**(10): 1758-1776

## 作者简介



钟 佳 (通信作者) 男, 1987 年出生, 博士, 中国科学院国家空间科学中心项目高级工程师, 主要研究方向为空间天气数据挖掘与机器学习建模、科学可视化技术。  
E-mail: zhongjia@nssc.ac.cn



邹自明 男, 1971 年出生, 博士, 中国科学院国家空间科学中心研究员、国家空间科学数据中心主任、中国科学院大学博士生导师, 长期从事空间科学与数据科学交叉领域研究。  
E-mail: mzou@nssc.ac.cn



徐寄遥 男, 1959 年出生, 博士, 中国科学院国家空间科学中心研究员、博士生导师, 国家杰出青年基金获得者, 现任国家重大科技基础设施“空间环境地基综合监测网——子午工程 II 期”总工程师。长期从事中高层大气物理学研究。  
E-mail: jyxu@spaceweather.ac.cn