0254-6124/2025/45(6)-1644-06 *Chin. J. Space Sci.* 空间科学学报

WANG Yitao, ZHANG Quanhao, LIU Jiajia. Dataset of Solar Prominences from 2011 to 2022 (in Chinese). *Chinese Journal of Space Science*, 2025, **45**(6): 1644–1649. CSTR:32142.14.cjss.2025–0089. DOI:10.11728/cjss2025.06.2025–0089

数据论文

2011 至 2022 年太阳日珥数据集

汪谊涛 [1] 张全浩 [1,2,3] 刘佳佳 [1,2,3]

1(中国科学技术大学地球和空间科学学院 合肥 230026)

2(中国科学技术大学比较行星学卓越创新中心 合肥 230026)

3(中国科学院近地空间环境重点实验室 合肥 230026)

摘 要 太阳日珥是悬浮在日冕中具有较低温度 (通常小于 10000 K) 和较稠密电子密度 (10^9～10^{11} cm^{-3}) 的磁结构. 相关研究表明, 太阳日珥与太阳耀斑及日冕物质抛射等可能引起灾害性空间天气的太阳爆发活动具有明显的关联性, 研究太阳日珥的时空分布有助于预报其空间天气效应, 从而尽可能降低太阳爆发活动带来的灾害性影响. 本数据集基于 10 min 分辨率的 SDO (Solar Dynamics Observatory) 卫星 AIA (Atmospheric Imaging Assembly) 仪器记录的 30.4 nm 数据图像, 通过背景重构增大日面外图像的对比度, 而后使用骨架提取、区域生长等自动算法标记重构后图像中的日珥区域, 获得相关参数值, 并将持续时间内存在的相同日珥追踪后存储于数据文件中, 最后将处理后的图像以及日珥数据文件存储在按年–月–日的三级结构目录中. 数据集囊括了从 2011 年 1 月 1 日 00:00 UT 至 2022 年 12 月 31 日 23:50 UT 共计 101741 个日珥文件, 严格按照相关协议和分类文件进行审核, 确保较高的可靠性, 为太阳日珥周期内的时空分布研究以及灾害性空间天气的预测提供科学数据支持.

关键词 太阳日珥, 空间天气, 太阳爆发活动, 太阳活动

中图分类号 P354

数据集基本信息

数据集名称	2011—2022年太阳日珥数据集
数据通信作者	张全浩 (zhangqh@ustc.edu.cn)
数据作者	汪谊涛, 张全浩, 刘佳佳
数据时间范围	2011年1月1日至2022年12月31日
数据空间范围	太阳色球层
数据分辨率	10 min
数据量	3.84 TByte
数据格式	*.jpg, *.txt
数据服务系统网址	DOI: https://doi.org/10.12176/01.01.00114-V01 CSTR: https://cstr.cn/CSTR:14804.11.01.01.00114-V01
基金项目	国家重点研发计划项目 (2022YFF0711400) , 中国科学院B类先导专项 (XDB0560000), 国家自然科学基金项目 (NSFC42188101, 42174213), 中国科学院网络安全与信息化专项项目 (CAS-WX2022SF-0103) 共同资助
数据库组成	数据库含2011年1月1日至2022年12月31日12年共计101741个追踪日珥, 包括参数数据文件及相关区域标记图像. 数据为 *.sav, *.txt和 *. jpg格式, 采用HGS空间坐标系

2025-06-06 收到原稿, 2025-09-19 收到修定稿

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

Dataset of Solar Prominences from 2011 to 2022

WANG Yitao[1] ZHANG Quanhao[1,2,3] LIU Jiajia[1,2,3]

1(*School of Earth and Space Sciences, University of Science and Technology of China, Hefei* 230026)

2(*Center for Excellence in Comparative Planetology, University of Science and Technology of China, Hefei* 230026)

3(*Key Laboratory of Geospace Environment, Chinese Academy of Sciences, Hefei* 230026)

Abstract Solar prominences are magnetic structures suspended in the corona, characterized by relatively low temperatures (typically below 10000K) and higher electron densities ($10^9 \sim 10^{11}$ cm^{-3}). Research indicates a clear correlation between prominences and solar eruptive activities, such as solar flares and coronal mass ejections that may trigger hazardous space weather. Studying the spatiotemporal distribution of solar prominences can aid in forecasting space weather efforts and help mitigate potential catastrophic impacts. This dataset is based on the 30.4 nm wavelength images captured by the Atmospheric Imaging Assembly (AIA) instrument aboard the Solar Dynamics Observatory (SDO) satellite, with a temporal resolution of 10 minutes. By employing background reconstruction to enhance the contrast of off-limb images, the automated algorithms, such as the skeleton extraction and the region-growing techniques, were used to identify prominence regions in the reconstructed images and extract relevant parameters. For those evolving in the same region during continuous frames, misidentification caused by duplicate naming is avoided by K-Nearest Neighbor (KNN) classification. Before tracking a procedure called non-prominence feature removal is used to discriminate real prominences from non-prominence features: Through Linear Discriminant Analysis (LDA), the eigenvalue of any target region can be calculated, and compare it with the derived distribution which is fitted with Gaussian distribution functions, to determine the likeliness of a real prominence, by which SLIPCAT can exclude active regions without involving other observation methods. Persisting prominences were tracked and stored in data files. At last, the processed images and prominence data files are organized in a year-month-day three-level directory structure. The dataset encompasses a total of 101741 prominence files, covering the period from 00:00 UT on 1 January 2011 to 23:50 UT on 31 December 2022. Rigorous validation was conducted in accordance with relevant protocols and classification standards to ensure high reliability. This dataset provides scientific support for research on the spatiotemporal distribution of solar prominences over their activity cycles and for the prediction of hazardous space weather events.

Key words Solar prominences, Space weather, Solar eruptive events, Solar events

Dataset Profile

Title	Dataset of Solar Prominences from 2011 to 2022
Data Corresponding Author	ZHANG Quanhao (zhangqh@ustc.edu.cn)
Data Authors	WANG Yitao, ZHANG Quanhao, LIU Jiajia
Time Range	1 Jan. 2011 to 31 Dec. 2022
Spatial Range	Solar chromosphere
Data Resolution	10 min
Data Volume	3.84 TByte
Data Format	*.jpg, *.txt
Data Service System URL	DOI: https://doi.org/10.12176/01.01.00114-V01 CSTR: https://cstr.cn/CSTR:14804.11.01.01.00114-V01
Sources of Funding	National Key Research and Development Program of China (2022YFF0711400), CAS Strategic Priority Research Program (XDB0560000), National Natural Science Foundation of China (NSFC42188101, 42174213), Cybersecurity and Informatization Special Project of Chinese Academy of Sciences (CAS-WX2022SF-0103)
Dataset Composition	This dataset contains a total of 101741 tracked prominences from 1 January 2011 to 31 December 2022, spanning 12 years. It includes parameter data files and corresponding region-labeled images. The data are stored in *.sav, *.txt, and *.jpg formats, using the Heliographic Stonyhurst (HGS) coordinate system

0 引言

太阳日珥是太阳上一类常见的结构, 广泛分布于太阳表面. 当观测日珥投影到日面上, 其被称作暗条. 作为从色球层向外延伸至日冕的巨大等离子体结构, 太阳日珥的温度一般比周围的日冕环境中的温度低一到两个数量级, 但是密度高一到两个数量级[1]. 在太阳磁场的作用下, 日珥形成弧状或环状结构. 日珥可分为宁静日珥和爆发日珥, 前者通常可稳定存在数周, 后者的生命在数小时到数天不等, 常伴随太阳耀斑或日冕物质抛射, 对地球空间环境产生显著影响[2].

近年来, 观测数据量日益庞大, 处理需求激增, 而研究者又需快速完成数据提取与挖掘, 因此有关太阳日珥/暗条的自动检测技术应运而生. 基于日珥的形态学与光学特征, 新兴的机器学习算法能够有效从大量日面边缘结构中精准识别日珥[3]. 例如, Bernasconi 等[4] 开发出一套能在 Hα 通道自动检测与追踪暗条的系统; Foullon 等[5] 通过对多波段极紫外通道数据进行比对, 实现了日面边缘日珥的自动检测, 并排除了非日珥结构; Loboda 团队[6] 利用 He II 304Å波段中宁静日珥的锐利边界特征, 开发了检测技术, 从 2008－2009 年 4 个月数据中识别出 389 个宁静日珥.

太阳临边日珥识别与追踪系统 (Solar Limb Prominence Catcher and Tracker, SLIPCAT) 最早由中国科学技术大学地球和空间科学学院汪毓明教授团队于 2010 年开发, 主要应用于日地关系观测台 (Solar Terrestrial Relations Observatory, STEREO) 产生的 EUVI (Extreme UltraViolet Imager) 仪器在 30.4 nm 波段数据[7], 可以从中自动识别和追踪太阳边缘的日珥结构. 2010 年 SDO (Solar Dynamics Observatory) 卫星成功发射升空, 其对太阳表面观测的时空精度相比 2006 年发射的 STEREO 卫星有大幅提升, 并且在接下来的十余年间, SDO 卫星很好地完成了多项任务, 并发回大量高质量的全日面观测图像. 本研究基于已有的 SLIPCAT 方法, 对其进行改进移植, 应用到 SDO 卫星 AIA (Atmospheric Imaging Assembly) 仪器获得的 30.4 nm 波段的观测数据中, 实现对 SDO 卫星高精度太阳观测数据的日珥自动识别与追踪. 对于具体的观测数据, 以 10 min 为间隔采样, 运用背景重构、骨架提取、区域生长等算法实现图像上日珥的识别标记与追踪. 实现一次识别标记过程一般为 50 s, 在所有图像的识别标记工作完成后方能进行追踪. 追踪过程是在多张图像中实现的, 追踪单张图像的所有日珥一般需要 2 s. 整个过程中生成的相关参数信息会存储在年–月–日三级结构目录中, 充足的数据量有助于在统计意义上考察一个周期内日珥空间分布的变化规律, 发掘其与同时间内耀斑及 CME 的关联, 为灾害性空间天气的预测研究提供了科学参考.

1 数据采集与处理方法

SDO 卫星的 AIA 仪器在 7 个极紫外波段和 3 个紫外–可见光波段拍摄太阳全日面图像, 分辨率高达 4096 pixel×4096 pixel (约 $0.6''\cdot\mathrm{pixel}^{-1}$), 每隔 12 s 完成一次全波段扫描, 并确保不同波段的图像严格对齐, 便于物理参数反演. 根据太阳实际活动强度, 自动调整曝光, 避免过饱和, 同时定期使用实验室标定数据修正望远镜的像场畸变和像素响应不均, 并通过冷却 CCD (–70°C 以下) 和数字滤波降低热噪声影响. 完成以上采集过程后, SDO 会通过高带宽 (130 $\mathrm{MByte\cdot s^{-1}}$) 的 Ka 波段天线, 每天向地面站传输约 1.5 TByte 数据.

本数据集基于 SDO/AIA 仪器传输的 30.4 nm 波段数据图像, 从 2011 年 1 月 1 日 00:00 UT 起, 每隔 10 min 采样一次, 至 2022 年 12 月 31 日 23:50 UT 止. 对于采样的每张图像, 在识别过程, 以背景重构增大日面外图像的对比度, 使边缘结构变得清晰可见, 再以骨架提取、区域生长等自动算法标记出日珥区域, 并获得相关参数值; 追踪前先以线性判别分析剔除非日珥特征结构: 采用高度、面积等参数, 基于划分的训练数据集建立线性判别方程并拟合, 从而根据方程对新检测的目标进行日珥/非日珥分类; 而后追踪过程中, 逐帧读取日珥参数数据文件, 将同一日珥的数据合并存储在新生成的追踪文件中, 追踪的日珥数据文件存储在第一帧出现的时间目录下.

2 数据样本描述

对于每个日珥, 生成的图像以 jpg 格式存储在按时间排布的目录下. 单帧图像中会标记日珥区域, 其对应日珥的参数文件通过指针形式存储在相应的数据文件中, 参数文件大小从几 kByte 到几千 kByte 不等, 与单帧图像中发现的日珥数量相关. 追踪的日珥

数据文件以指针形式存储在相应的数据文件中, 并且标注在 jpg 图像上, 其文件大小从几 kByte 到几千 kByte 不等, 与追踪的日珥的生命周期长度相关. 日珥参数信息列于表 1.

图 1 给出了带有标记的日珥区域及参数的图像, 获得的日珥各项参数信息的描述列于表 1. 需要注意的是两个比较复杂的参数L_{lumpy}和E_{elong}, 分别定义为

$$L_{\text{lumpy}} = \frac{4 \times A_{\text{area}}}{\pi \max\left[\left(x_{1,\text{p}} - x_{2,\text{p}}\right)^2 + \left(y_{1,\text{p}} - y_{2,\text{p}}\right)^2\right]},$$

$$E_{\text{elong}} = \frac{S_{\text{spine_len}}}{A_{\text{area}}}.$$

其中, $x_{1,\text{p}}, y_{1,\text{p}}, x_{2,\text{p}}, y_{2,\text{p}}$分别为日珥区域内不同两个像素点的横纵坐标, A_{area}与$S_{\text{spine_len}}$分别为区域面积与中脊线长度, 由程序内置算法经过复杂计算得出.

3 数据质量控制

SDO 卫星发射迄今已有 15 年, 传回的图片精度相当于普通高清电视的数倍, 从其下载得到的观测数据经检查像素完整, 表现出较高的质量水平, 因此可直接读取并进行处理[8]. 在识别阶段, 采用多次多组数据进行训练以确定模型中的超参数设置, 进而确保日珥的识别率在一个较高的范围. 同时考虑到噪声, SLIPCAT 采用多种图像算法尽可能地剔除噪声影响, 以更精准地识别出日珥, 并提取真实的参数数据. 对于识别过程中存在的非日珥结构, SLIPCAT 采用线性判别分析法[9] 剔除非日珥的结构. 结果表明, 在不同组一定时间内的测试中 SLIPCAT 均能表现出较高的识别率, 同时, 在太阳周期内, 不同时期的识别率相差不大, 提取出的日珥信息符合实际情况, 保证了数据质量的可靠性.

4 数据价值

数据集收录了原始 SDO 全日面图像、背景重构后的全日面图像、标记出日珥区域的全日面图像以及标记了日珥区域和部分参数的全日面图像, 同时还收录了每帧日珥检测到的日珥参数信息文件以及追踪

表 1 日珥参数信息描述

Table 1 Prominence parameter details

序号	参数	信息描述
1	SID	日珥 id
2	BOUNDARIES	指针型变量, 包含标记的日珥边界坐标(x, y, 单位pixel) 以及默认设定的算法阈值 (info)
3	COM	包含两个元素的一维数组, 加权后的日心距(单位为1R_{s}, R_{s}表示太阳半径)和方位角(单位为°)
4	SPA	日珥的所有像素点的最低和最高方位角 (单位为°)
5	SPR	日珥的所有像素点的最低高度和最大高度, 单位为1R_{s}
6	AREA	日珥的面积, 单位 Mm^2
7	PERIMETER	日珥的周长, 单位 Mm
8	FLUX	日珥的总亮度, 单位 DN (Digital Number)
9	SPINE-POS	日珥中脊线 (骨架主干线) 长度, 单位 Mm
10	SPINE-LEN	日珥的骨架参数信息, 包含两个参数: skel_info为默认设定的算法阈值, skel_ xy为骨架的每个坐标点的信息, 单位 pixel
11	FOOT	日珥的足点数量
12	CAVE	未定义参数, 可以忽略
13	FVAR	包含4个元素的一维数组, LDA线性拟合过程中的结果参数
14	LUMPY	日珥的形态学参数定义为日珥面积与日珥区域像素坐标最远距离为直径的圆面积比值. 如果比值接近1, 说明日珥的区域形状接近正圆
15	ELONG	日珥的形态学参数定义为中脊线长度与日珥面积的比值. 比值越接近1, 说明日珥长宽比越大; 比值越接近0, 说明日珥越接近正圆
16	CFL	日珥的置信度, 非1代表非日珥, 1代表日珥
17	MISSING	中间参数可以忽略

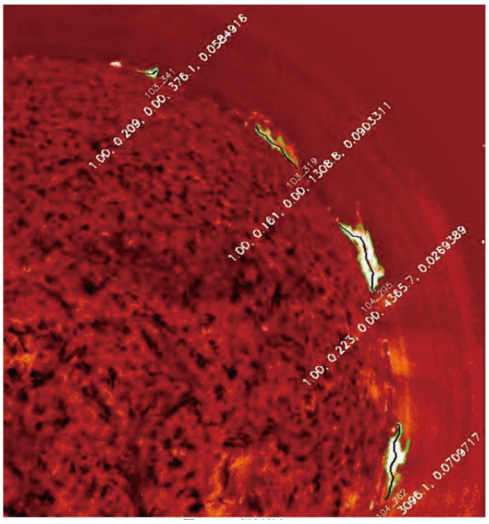

图 1 日珥样例的标记

Fig. 1 Labeling of prominence samples

完成的日珥参数信息文件. 2013 年 11 月 2 日 23:40 UT 的单帧处理结果如图 2 所示, 由于两个参数信息繁多, 仅以单个日珥样例的参数信息展示在图 3 中. 由于数据量庞大, 边界坐标信息 BOUNDARIES 和中脊线坐标信息 SPINE_POS 以指针形式存储. 需要注意的是, 追踪完成的日珥参数信息是对不同帧图像中检测到的同一日珥的信息汇总, 因此对于某一特定时刻的日珥信息, 两个参数文件中存储信息一致.

该数据集支持对太阳长周期内日珥的空间分布情况及演化进行综合分析, 而且能够基于研究 结果探究其中的爆发日珥与同时期内发生的耀斑、日冕物质抛射之间的关联性, 为灾害性空间天气的预报工作提供支持[10].

致谢 国家科技基础条件平台——国家空间科学数据中心 (http://www.nssdc.ac.cn) 为本文提供了数据资源支持.

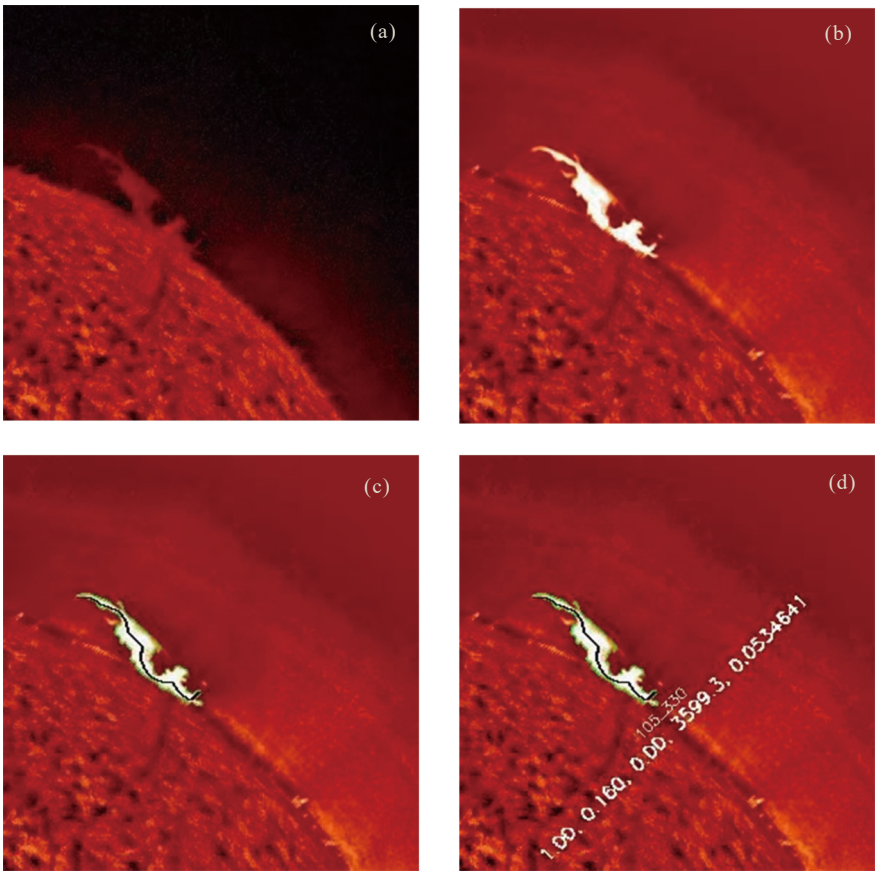

图 2 2013 年 11 月 2 日 23:40 UT 的单帧图像处理结果. (a) 原始图像, (b) 背景重构后的全日面图像, (c) 标记出日珥区域的全日面图像, (d) 标记出日珥区域及部分参数的全日面图像

Fig. 2 Processing results of the single frame at 23:40 UT on 2 November 2013. (a) Original image, (b) full-disk image after background reconstruction, (c) full-disk image with prominence regions marked, (d) full-disk image with prominence regions and partial parameters marked

(a)
```
"SID": "20131030_211007_SDO_104_336.1.1.1",
"BOUNDARIES": [<PtrHeapVar282>, <PtrHeapVar283>],
"COM": [1.0994862, 337.59860],
"SPA": [335.77158, 339.48669],
"SPR": [1.0175780, 1.2176182],
"AREA": 3050.2371,
"PERIMETER": 386.75818,
"FLUX": 16.685385,
"SPINE_POS": [<PtrHeapVar284>, <PtrHeapVar285>],
"SPINE_LEN": 144.69670,
"FOOT": 1,
"CAVE": 0,
"FVAR": [-5.7892914, 0.50260341, 1.0000000,
                                 5.0000000],
"LUMPY": 0.21024910,
"ELONG": 0.047437854,
"CFL": 1.0000000,
"MISSING": 0
```
(b)
```
"SID": "20131030_211007_SDO_104_336.1.1.1",
"BOUNDARIES": [<PtrHeapVar282>, <PtrHeapVar283>],
"COM": [1.0994862, 337.59860],
"SPA": [335.77158, 339.48669],
"SPR": [1.0175780, 1.2176182],
"AREA": 3050.2371,
"PERIMETER": 386.75818,
"FLUX": 16.685385,
"SPINE_POS": [<PtrHeapVar284>, <PtrHeapVar285>],
"SPINE_LEN": 144.69670,
"FOOT": 1,
"CAVE": 0,
"FVAR": [-5.7892914, 0.50260341, 1.0000000,
                                 5.0000000],
"LUMPY": 0.21024910,
"ELONG": 0.047437854,
"CFL": 1.0000000,
"MISSING": 0
```


图 3 单个日珥样例参数信息. (a) 2013 年 11 月 2 日 23:40 UT 分日珥参数信息, (b) 日珥追踪过程存储参数信息

Fig. 3 Parameter information of a single prominence sample. (a) Prominence parameters at 23:40 UT on 2 November 2013, (b) parameters stored during the prominence tracking process

参考文献

[1] LABROSSE N, HEINZEL P, VIAL J C, *et al.* Physics of solar prominences: I—spectral diagnostics and non-LTE modelling[J]. *Space Science Reviews*, 2010, **151**(4): 243-332

[2] KOLEVA K, GOPALSWAMY N, DEVI P, *et al.* Spatial relationship between CMEs and prominence eruptions during SC 24 and SC 25[J]. *The Astrophysical Journal*, 2024, **966**(1): 22

[3] FILIPPOV B P, DEN O G. Prominence height and vertical gradient in magnetic field[J]. *Astronomy Letters*, 2000, **26**(5): 322-327

[4] BERNASCONI P N, RUST D M, HAKIM D. Advanced automated solar filament detection and characterization code: description, performance, and results[J]. *Solar Physics*, 2005, **228**(1/2): 97-117

[5] FOULLON C, VERWICHTE E. Automated detection of EUV prominences[J]. *Solar Physics*, 2006, **234**(1): 135-150

[6] LOBODA I P, BOGACHEV S A. Quiescent and eruptive prominences at solar minimum: a statistical study via an automated tracking system[J]. *Solar Physics*, 2015, **290**(7): 1963-1980

[7] WANG Y M, CAO H, CHEN J H, *et al.* Solar Limb Prominence Catcher and Tracker (SLIPCAT): an automated system and its preliminary statistical results[J]. *The Astrophysical Journal*, 2010, **717**(2): 973-986

[8] LEMEN J R, TITLE A M, AKIN D J, *et al.* The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO)[J]. *Solar Physics*, 2012, **275**(1/2): 17-40

[9] FISHER R A. The use of multiple measurements in taxonomic problems[J]. *Annals of Human Genetics*, 1936, **7**(2): 179-188

[10] PARENTI S. Solar prominences: observations[J]. *Living Reviews in Solar Physics*, 2014, **11**(1): 1

作者简介

汪谊涛（第一作者） 男, 2000 年 7 月出生于安徽省安庆市, 现为中国科学技术大学空间科学与物理学院硕士研究生, 主要研究方向为太阳日珥事件过程相关研究.
E-mail: 484247926@qq.com

张全浩（通信作者） 男, 2017 年毕业于中国科学技术大学, 获博士学位, 现为中国科学技术大学地球和空间科学学院副教授, 主要研究方向为利用观测数据分析和磁流体力学数值模拟, 研究太阳和行星际空间中的活动现象.
E-mail: zhangqh@ustc.edu.cn