中国科学院国家天文台
NATIONAL ASTRONOMICAL OBSERVATORIES, CAS

National Astronomical Data Center
国家天文科学数据中心

China-VO

XPhotonbank: AI-Ready的光子数据库

报 告 人：张震
团队成员：崔辰州，许允飞、袁为民、刘元、李东悦等
单　　位：国家天文台、国家天文科学数据中心

Contents

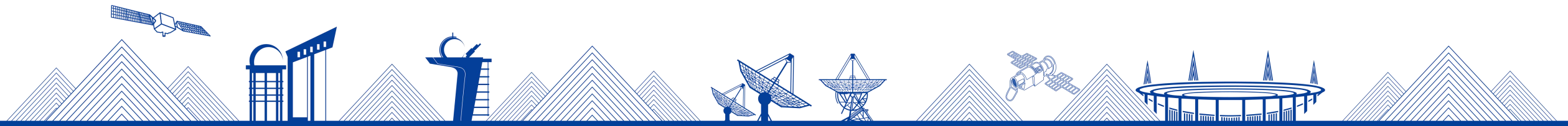
National Astronomical Data Center
国家天文科学数据中心

01 EP及其数据处理流水线

02 光子数据潜在价值

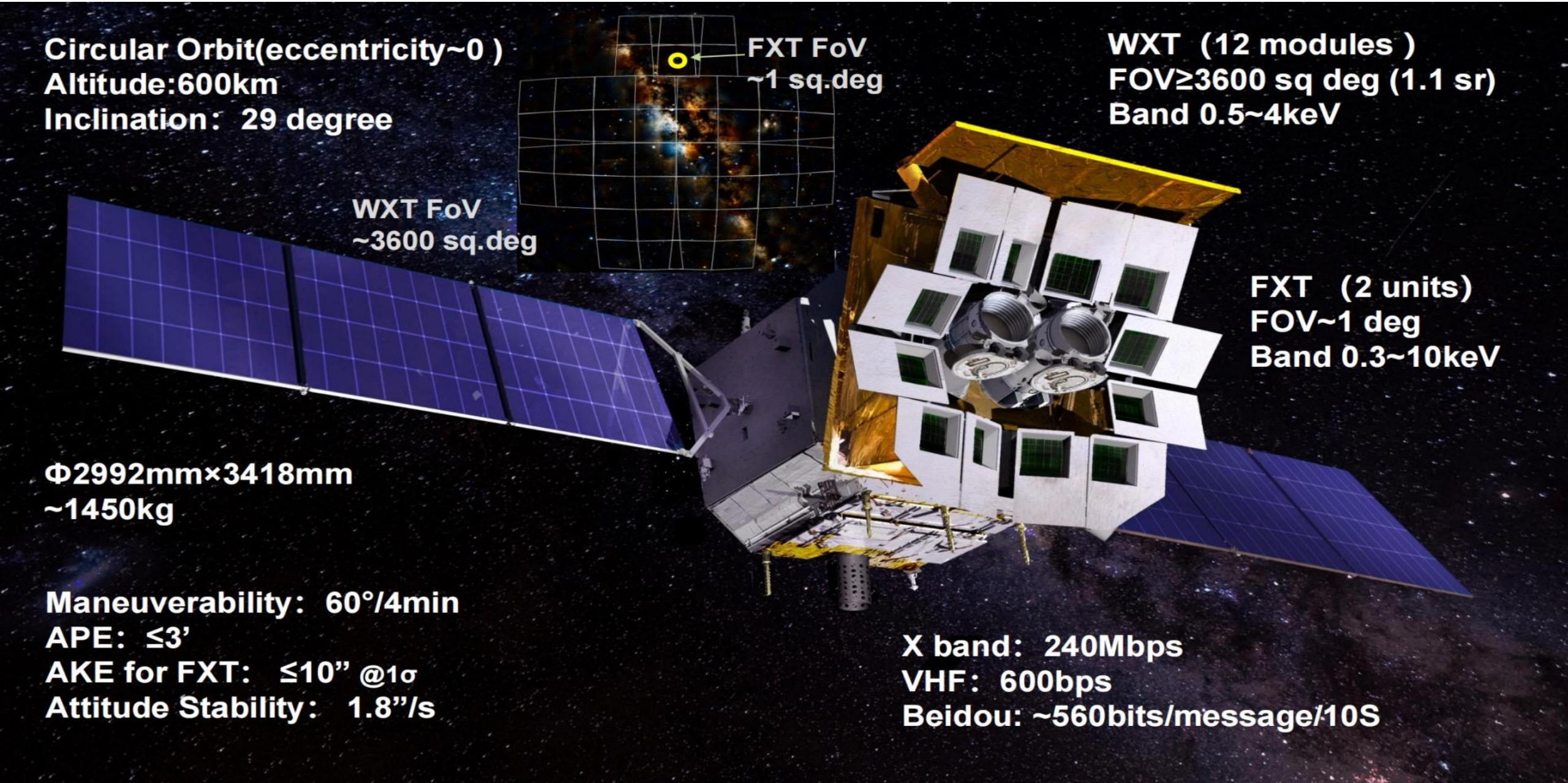
03 光子数据库的构建

04 应用场景与未来计划



天关卫星

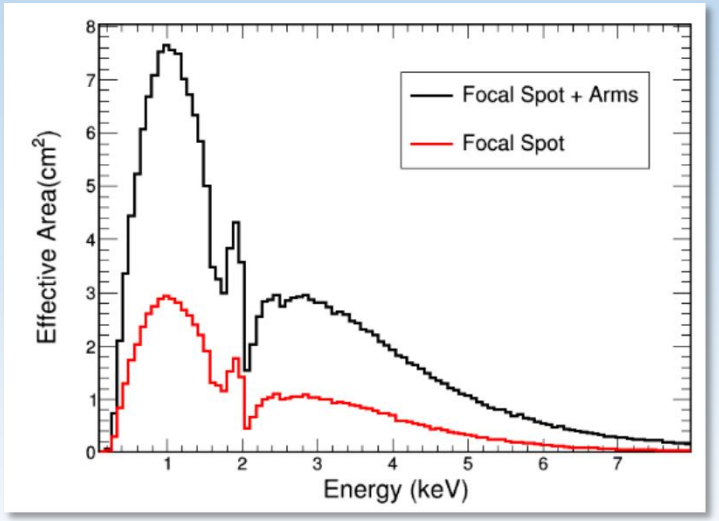
National Astronomical Data Center
国家天文科学数据中心



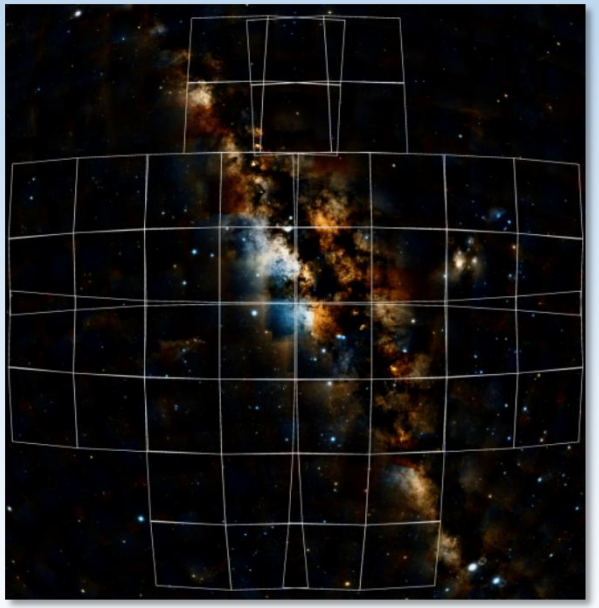
天关卫星的核心优势

National Astronomical Data Center
国家天文科学数据中心

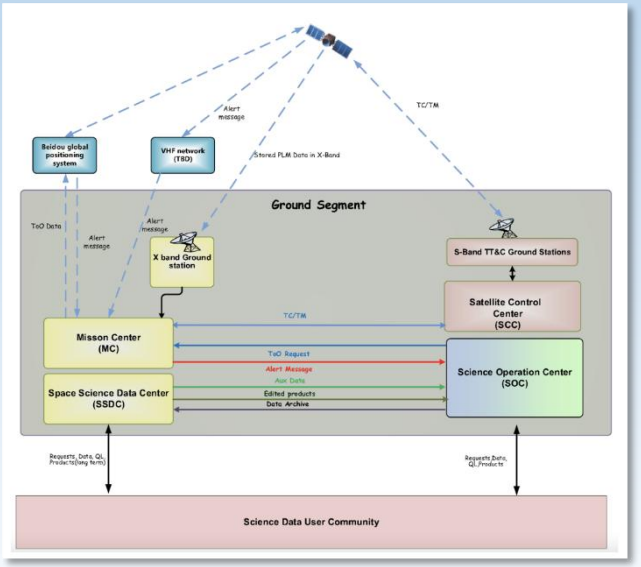
软X波段的高灵敏度



大视场

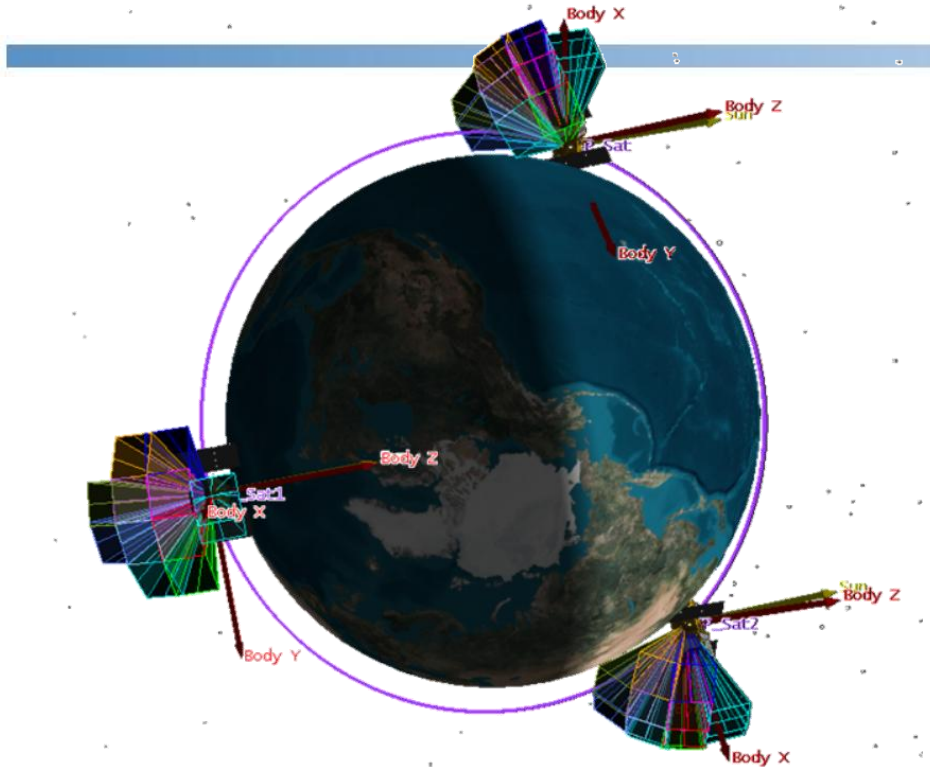


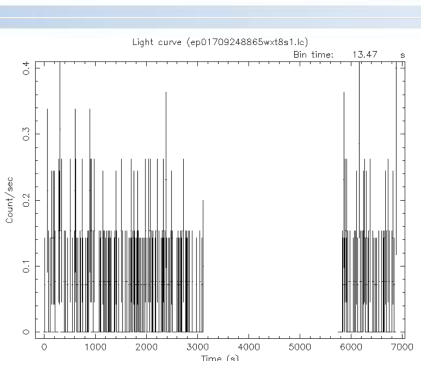
快速后随

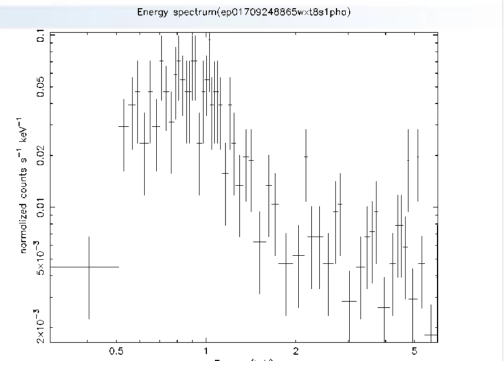


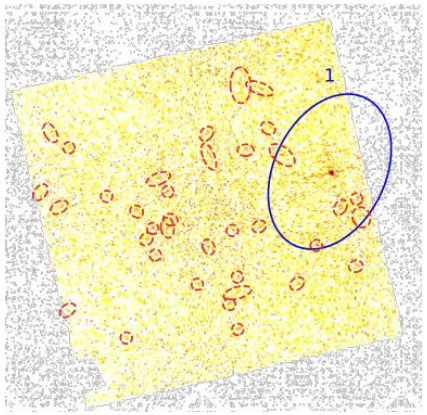
流水线数据处理单元: OBS_ID

National Astronomical Data Center
国家天文科学数据中心









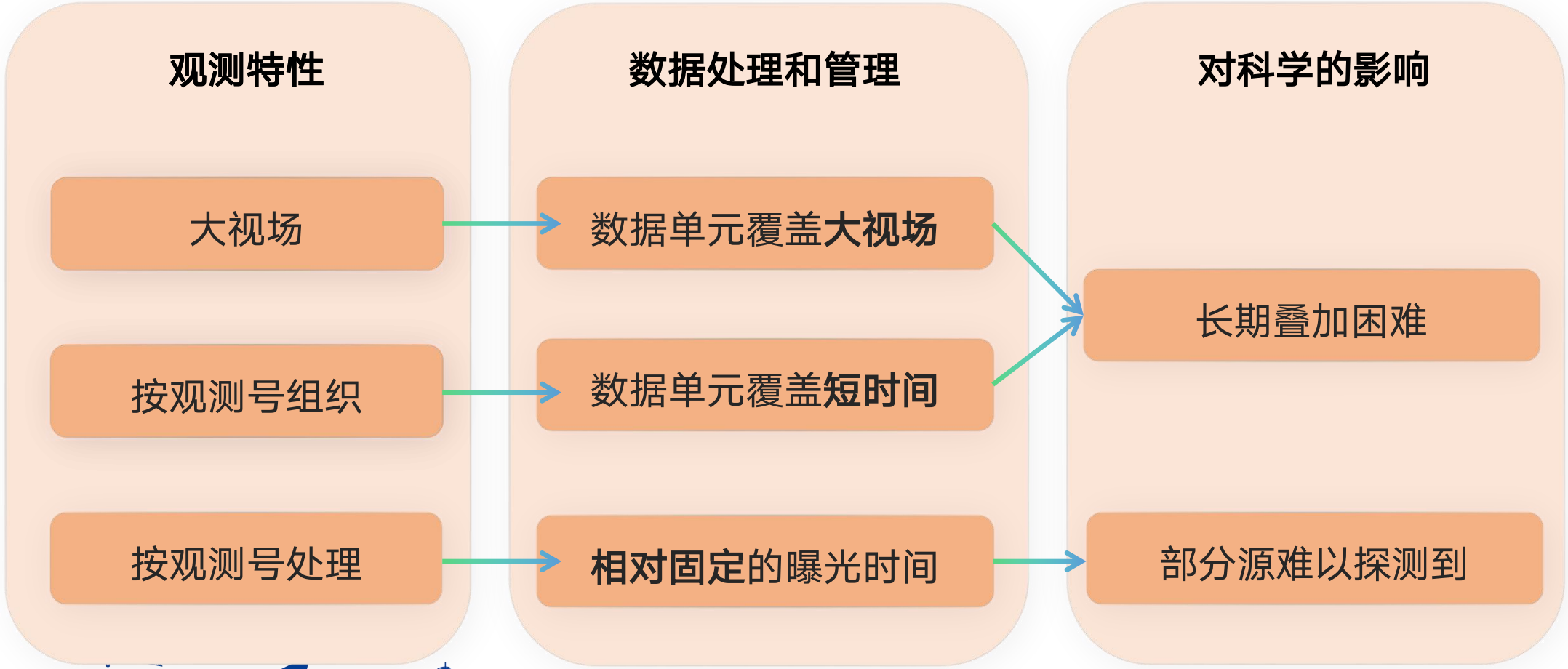
ep01709248865wxt9s1	2025-11-26 13:54:57	54.312	25.982
ep01709248865wxt24s3	2025-11-26 13:54:57	8.971	59.822
ep01709248865wxt24s1	2025-11-26 13:54:57	6.320	64.155
ep01709248865wxt24s2	2025-11-26 13:54:57	14.191	60.721
ep01709248865wxt9s2	2025-11-26 13:54:57	58.823	31.078
ep01709248865wxt21s1	2025-11-26 13:54:43	32.379	52.425
ep01709248865wxt7s2	2025-11-26 13:54:35	47.040	40.951
ep01709248865wxt7s1	2025-11-26 13:54:35	49.958	41.512
ep01709248865wxt2s1	2025-11-26 13:54:13	44.288	53.894
ep01709248865wxt22s1	2025-11-26 13:53:34	26.611	61.751
ep01709248865wxt35s1	2025-11-26 13:53:21	63.361	10.492
ep01709248865wxt35s2	2025-11-26 13:53:21	57.676	5.792

一轨三次测的Monitoring Survey，每次观测平均曝光1000s，一次观测作为一个数据处理单元

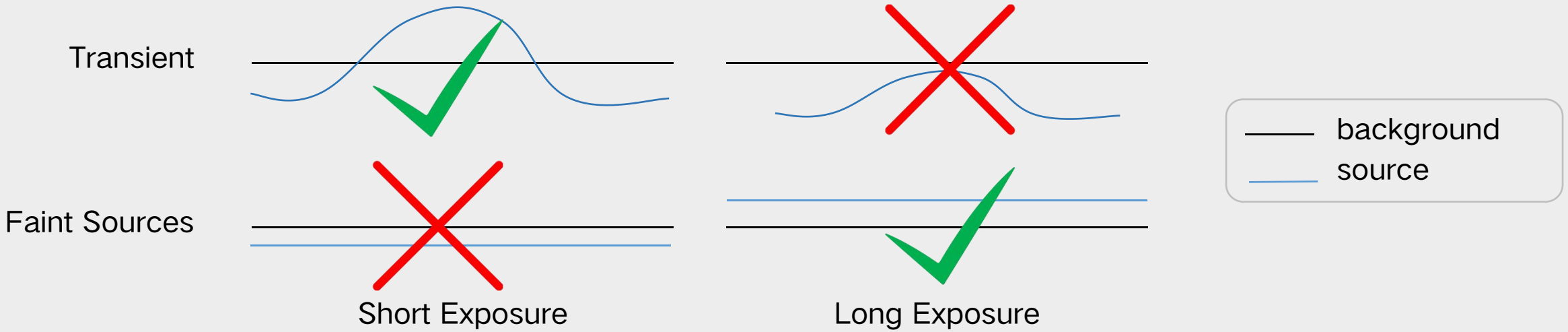
每次观测产生的数据产品：光变、能谱、图像、星表

现有数据处理方式局限性

National Astronomical Data Center
国家天文科学数据中心



对源探测的影响



固定的曝光时间可能漏掉部分源

对长期演化研究的影响

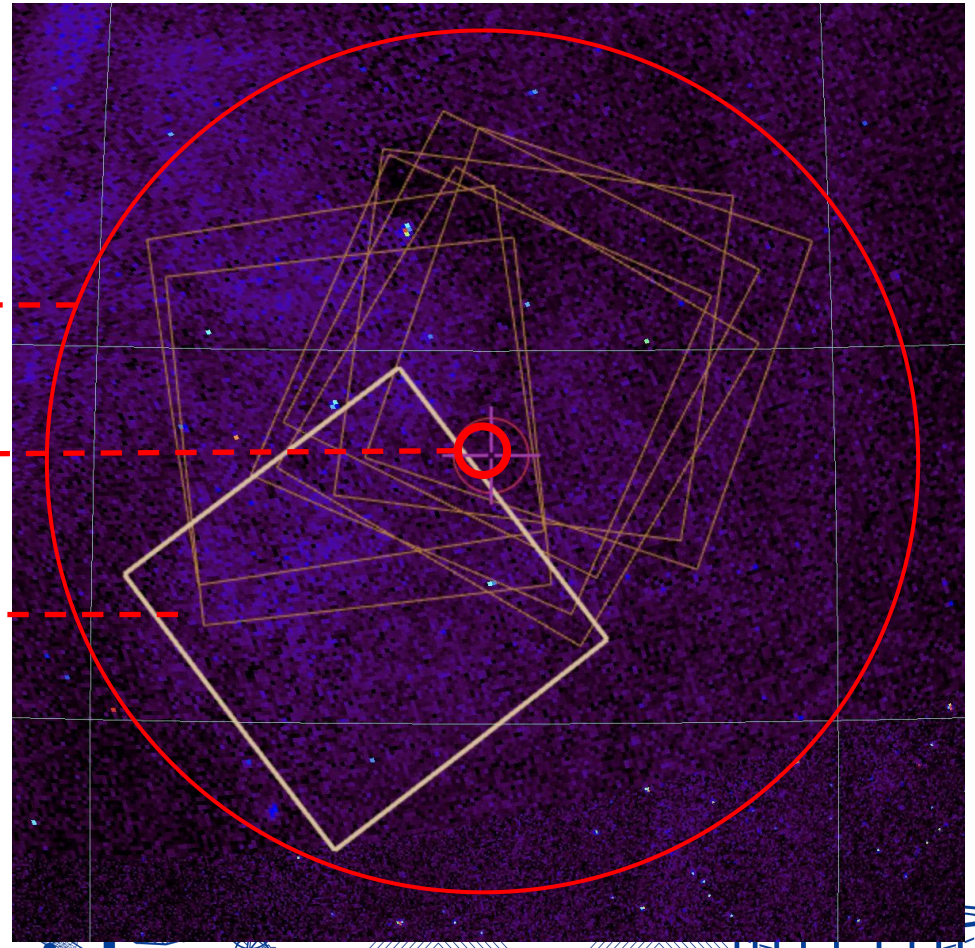
National Astronomical Data Center
国家天文科学数据中心

只有1/200的数据是需要的

检索到的：25° x 25°

需要的：1° x 1°

CMOS FOV：9° x 9°



Contents

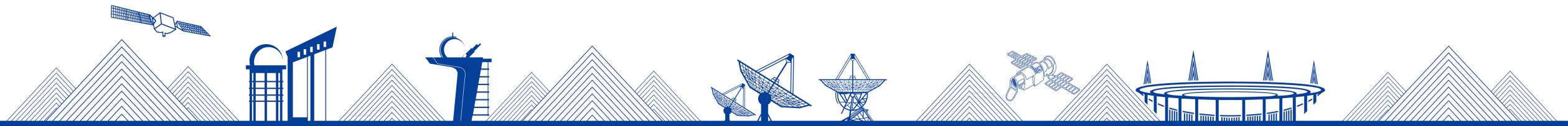
National Astronomical Data Center
国家天文科学数据中心

01 EP及其数据处理流水线

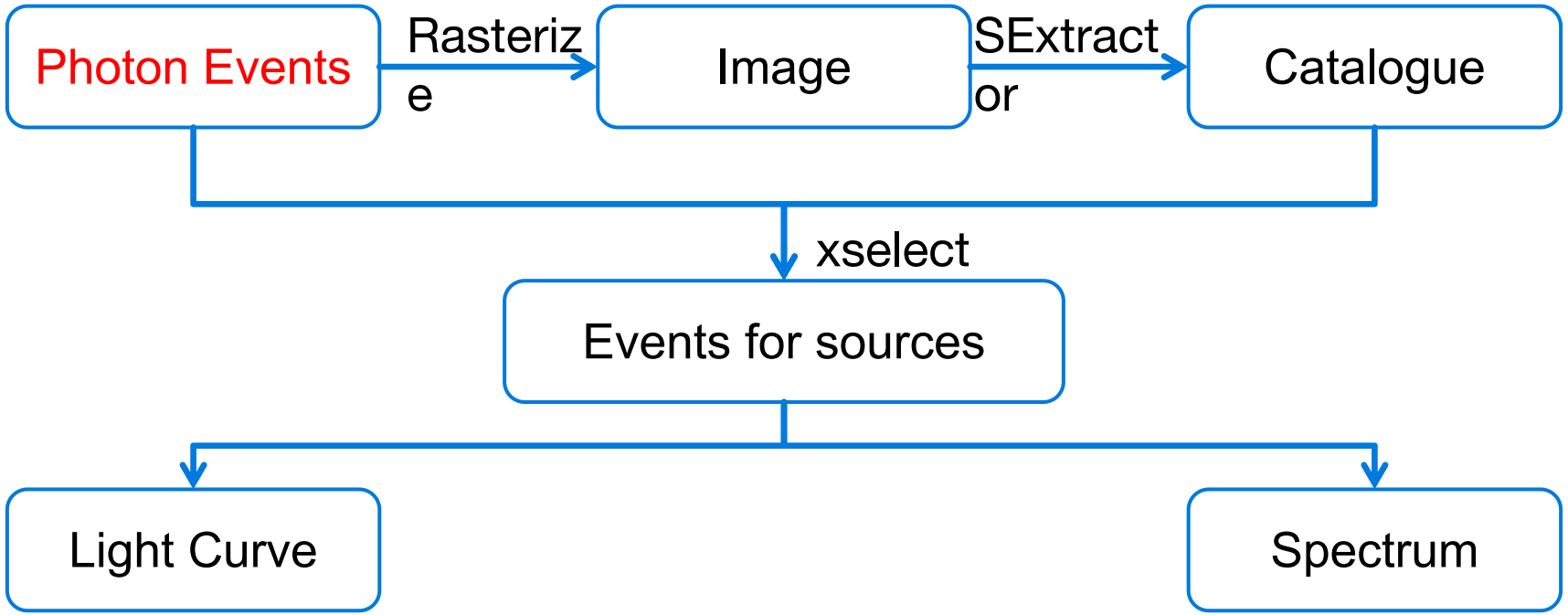
02 光子数据潜在价值

03 光子数据库的构建

04 应用场景与未来计划

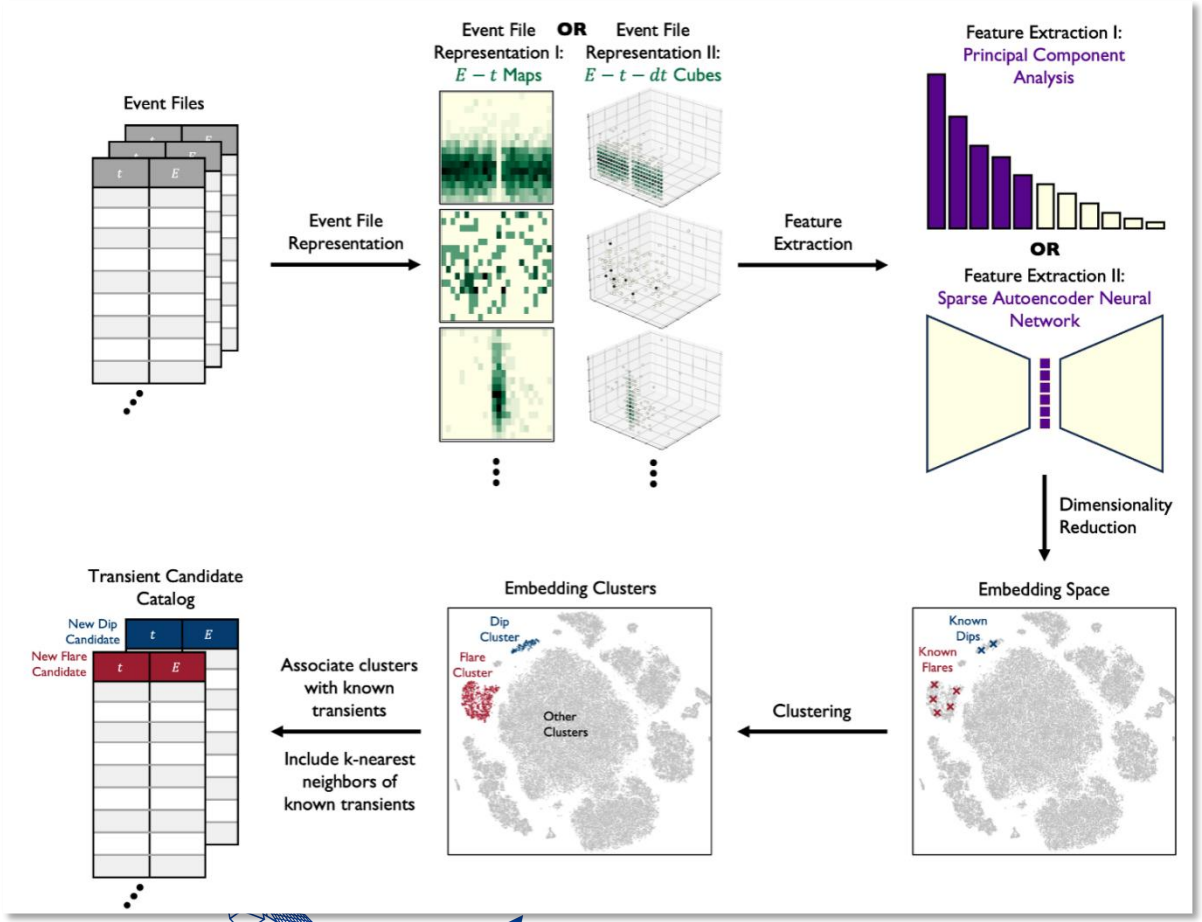


光子数据



光子数据是其他科学产品的来源

光子数据



从数据科学视角，光子数据作为更高维的信息，可能蕴含更多科学价值。 (Dillmann et al., 2025)

其他高能任务的光子数据库

National Astronomical Data Center
国家天文科学数据中心

Currently Available Data Products

The Fermi data released to the scientific community is governed by the data policy. The released instrument data for the GBM, along with LAT source lists, can be accessed through the Browse interface specific to Fermi. LAT photon data can be accessed through the LAT data server.

The FITS files can also be downloaded from the Fermi FTP site. The file version number is the 'xx' in the characters before the extension in each filename; you should keep track of the version numbers of files you analyze since the instrument teams may update them.

Note that the LAT and GBM data are accompanied by caveats about their use.

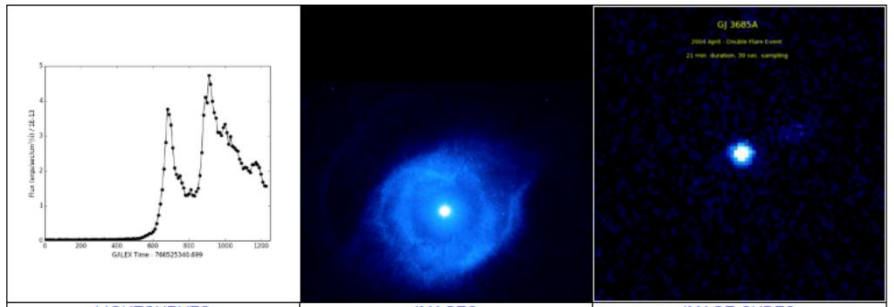
- LAT Photon and Extended Data
  - LAT Data Server (updated with P8R3 data 26-Nov-2018)
  - LAT Low-Energy (LLE) Data (Browse table)
  - Products available on the FTP Site (current processing version of the data).
    - Weekly Photon Files
    - Weekly Spacecraft Files
    - Mission Long Spacecraft File
    - Weekly 1-second Spacecraft Files
    - Filtered Weekly Photon Files with Diffuse Response Columns
  - Previous processing versions available on the FTP site
    - Pass 8 (P8R2) Weekly Files
    - Pass 7 (V6d) Weekly Files
    - Pass 7 (V6) Weekly Files
    - Pass 6 (V11) Weekly Files
    - Pass 6 (V3) Weekly Files
  - SSDC data server (external)
- LAT catalogs and associated products (high-level products only)
  - LAT Source Catalog
    - LAT 14-year Source Catalog (4FGL-DR4)
    - LAT 12-year Source Catalog (4FGL-DR3)
    - LAT 10-year Source Catalog (4FGL-DR2)
    - LAT 8-year Source Catalog (4FGL)
    - Preliminary LAT 8-year Source List (FL8Y)
    - LAT 4-year Source Catalog (3FGL)

■ Weekly Photon Files

Fermi Data Repository

Introduction

There are three primary data products that can be created using gPhoton.



LIGHTCURVES | IMAGES | IMAGE CUBES

There are three primary Python modules you will use.

1. **gFind** - The search tool. Queries the database for available data (reported as time ranges) at a given coordinate.
2. **gAperture** - The photometry tool. Performs aperture photometry at the given coordinate and returns light curves at a specified bin width. Can optionally write a "photon event file" containing each event.
3. **gMap** - The image tool. Writes FITS images or image cubes (image frames) of a requested sky region.

Background

The Galaxy Evolution Explorer (GALEX) mission surveyed the sky in the ultraviolet for approximately ten years between April 2003 and June 2013. Its two microchannel plate (MCP) detectors recorded the position and time-of-arrival of incident photons ("photon events") with a time resolution of 5 thousandths of a second. Due to storage space and processing limitations, however, these data have previously been released and archived almost exclusively as integrated images (with visit- and coadd-level depths of minutes or more). The gPhoton database and associated software tools now give researchers the means to use photon-level GALEX data up through General Release (GR) 6/7.

The corpus of GR 6/7 photon data is composed of more than 79 billion FUV and over a trillion NUV photon events. On disk, the gPhoton database is approximately 130 TB. The gPhoton command line tools permit users to search for coverage within this database and create calibrated light curves, images, or movies, with the flexibility to set time and spatial ranges or cuts across the entire mission. This enables studies of variability at timescales shorter than the GALEX visit files, customization or refinement of the calibration, and simplifies the fusion of data across mission phases, surveys and modes. When imported as Python modules, gPhoton permits even more detailed work with the data, including analysis at the level of individual photons.

GALEXY gPhoton

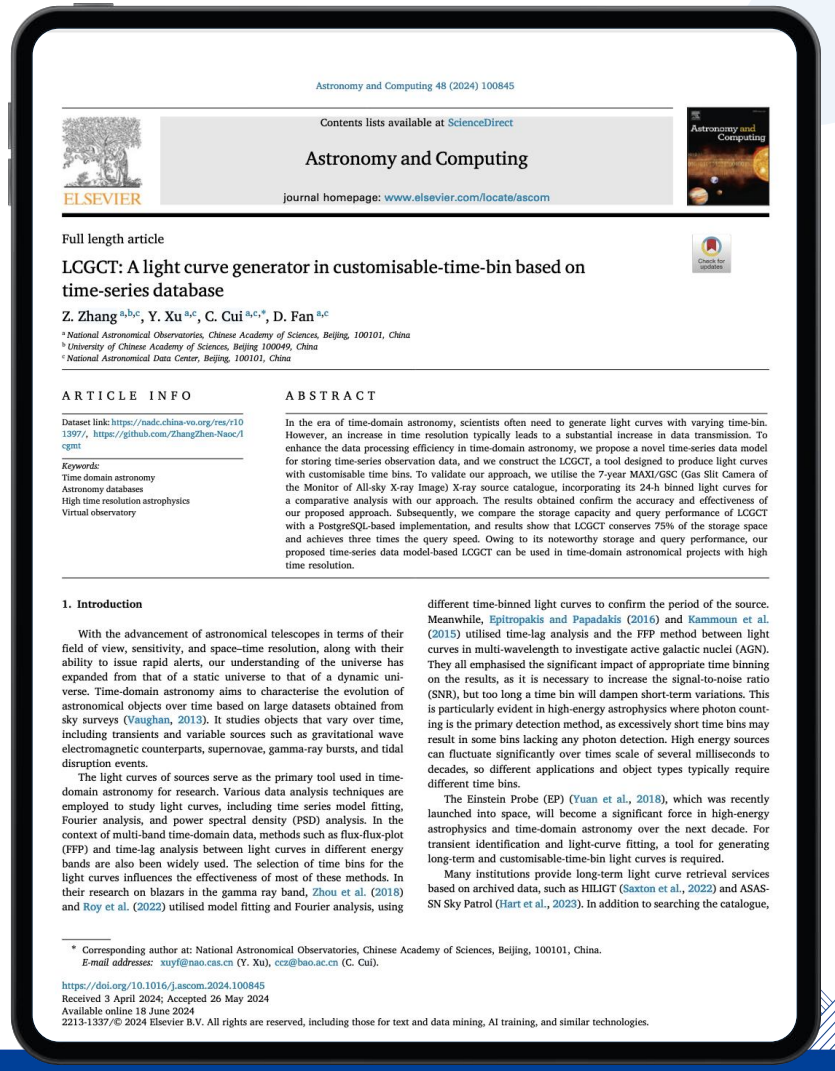
相关工作: NADC LCGCT

National Astronomical Data Center
国家天文科学数据中心

Purpose: Generate lightcurve on customisable time-bin

Technology: Time-series database

Application Area: Transient / Variable source study in high cadence obsrevation



Astronomy and Computing 48 (2024) 100845

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

LCGCT: A light curve generator in customisable-time-bin based on time-series database

Z. Zhang [a,b,c], Y. Xu [a,c], C. Cui [a,c,*], D. Fan [a,c]

[a] National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] National Astronomical Data Center, Beijing, 100101, China

ARTICLE INFO

Dataset link: https://nadc.china-vo.org/res/r101397/, https://github.com/ZhangZhen-Naoc/lcgmt

Keywords:
Time domain astronomy
Astronomy databases
High time resolution astrophysics
Virtual observatory

ABSTRACT

In the era of time-domain astronomy, scientists often need to generate light curves with varying time-bin. However, an increase in time resolution typically leads to a substantial increase in data transmission. To enhance the data processing efficiency in time-domain astronomy, we propose a novel time-series data model for storing time-series observation data, and we construct the LCGCT, a tool designed to produce light curves with customisable time bins. To validate our approach, we utilise the 7-year MAXI/GSC (Gas Slit Camera of the Monitor of All-sky X-ray Image) X-ray source catalogue, incorporating its 24-h binned light curves for a comparative analysis with our approach. The results obtained confirm the accuracy and effectiveness of our proposed approach. Subsequently, we compare the storage capacity and query performance of LCGCT with a PostgreSQL-based implementation, and results show that LCGCT conserves 75% of the storage space and achieves three times the query speed. Owing to its noteworthy storage and query performance, our proposed time-series data model-based LCGCT can be used in time-domain astronomical projects with high time resolution.

1. Introduction

With the advancement of astronomical telescopes in terms of their field of view, sensitivity, and space–time resolution, along with their ability to issue rapid alerts, our understanding of the universe has expanded from that of a static universe to that of a dynamic universe. Time-domain astronomy aims to characterise the evolution of astronomical objects over time based on large datasets obtained from sky surveys (Vaughan, 2013). It studies objects that vary over time, including transients and variable sources such as gravitational wave electromagnetic counterparts, supernovae, gamma-ray bursts, and tidal disruption events.

The light curves of sources serve as the primary tool used in time-domain astronomy for research. Various data analysis techniques are employed to study light curves, including time series model fitting, Fourier analysis, and power spectral density (PSD) analysis. In the context of multi-band time-domain data, methods such as flux-flux-plot (FFP) and time-lag analysis between light curves in different energy bands are also been widely used. The selection of time bins for the light curves influences the effectiveness of most of these methods. In their research on blazars in the gamma ray band, Zhou et al. (2018) and Roy et al. (2022) utilised model fitting and Fourier analysis, using different time-binned light curves to confirm the period of the source. Meanwhile, Epitropakis and Papadakis (2016) and Kammoun et al. (2015) utilised time-lag analysis and the FFP method between light curves in multi-wavelength to investigate active galactic nuclei (AGN). They all emphasised the significant impact of appropriate time binning on the results, as it is necessary to increase the signal-to-noise ratio (SNR), but too long a time bin will dampen short-term variations. This is particularly evident in high-energy astrophysics where photon counting is the primary detection method, as excessively short time bins may result in some bins lacking any photon detection. High energy sources can fluctuate significantly over times scale of several milliseconds to decades, so different applications and object types typically require different time bins.

The Einstein Probe (EP) (Yuan et al., 2018), which was recently launched into space, will become a significant force in high-energy astrophysics and time-domain astronomy over the next decade. For transient identification and light-curve fitting, a tool for generating long-term and customisable-time-bin light curves is required.

Many institutions provide long-term light curve retrieval services based on archived data, such as HILIGT (Saxton et al., 2022) and ASAS-SN Sky Patrol (Hart et al., 2023). In addition to searching the catalogue,

* Corresponding author at: National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100101, China.
E-mail addresses: xuyf@nao.cas.cn (Y. Xu), ccz@bao.ac.cn (C. Cui).

https://doi.org/10.1016/j.ascom.2024.100845
Received 3 April 2024; Accepted 26 May 2024
Available online 18 June 2024
2213-1337/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Contents

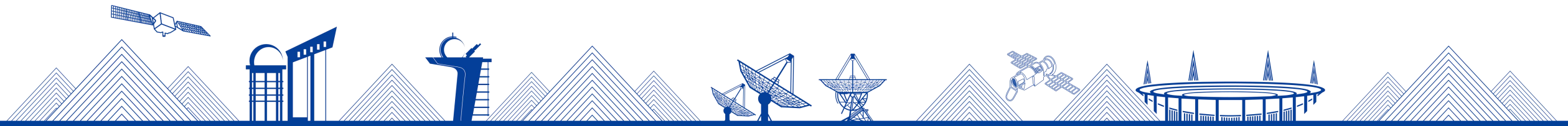
National Astronomical Data Center
国家天文科学数据中心

01 EP及其数据处理流水线

02 光子数据潜在价值

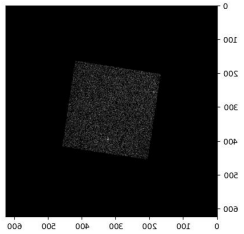
03 光子数据库的构建

04 应用场景与未来计划

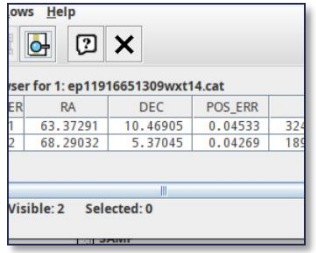


元数据: Arm Factor

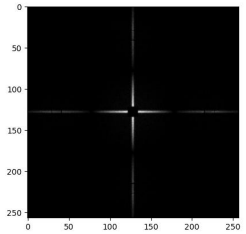
National Astronomical Data Center
国家天文科学数据中心



Original Image



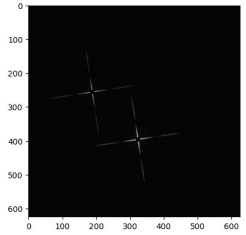
Catalog



PSF

具体光子是否来自Arm？

连续光子来自Arm的比例？



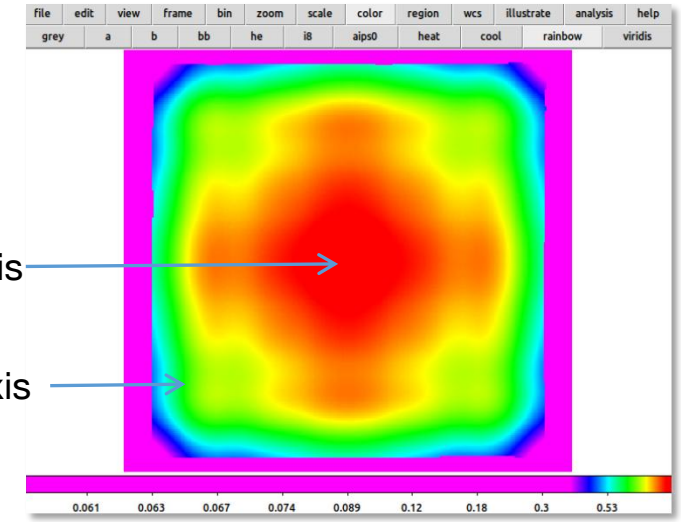
Model Image

Original Image

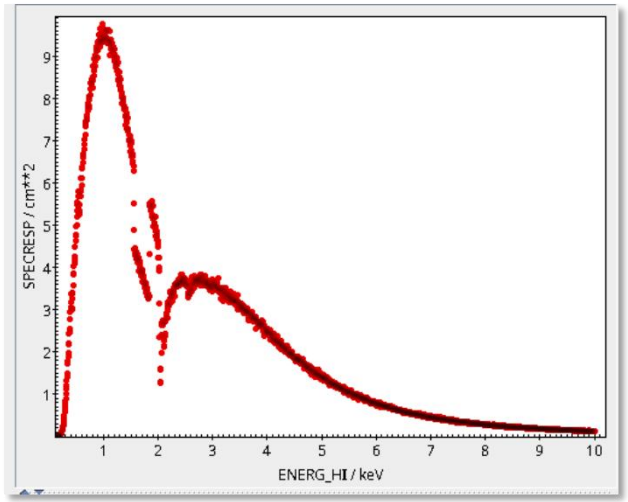
Arm Ratio

数据校准: Scale Factor

National Astronomical Data Center
国家天文科学数据中心



×



不同像素位置的有效面积

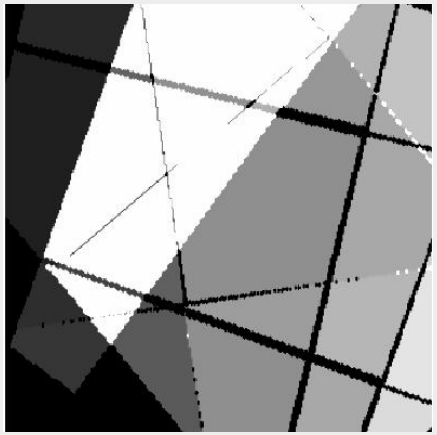
不同能量的有效面积

Scale Factor

数据对齐: GTI



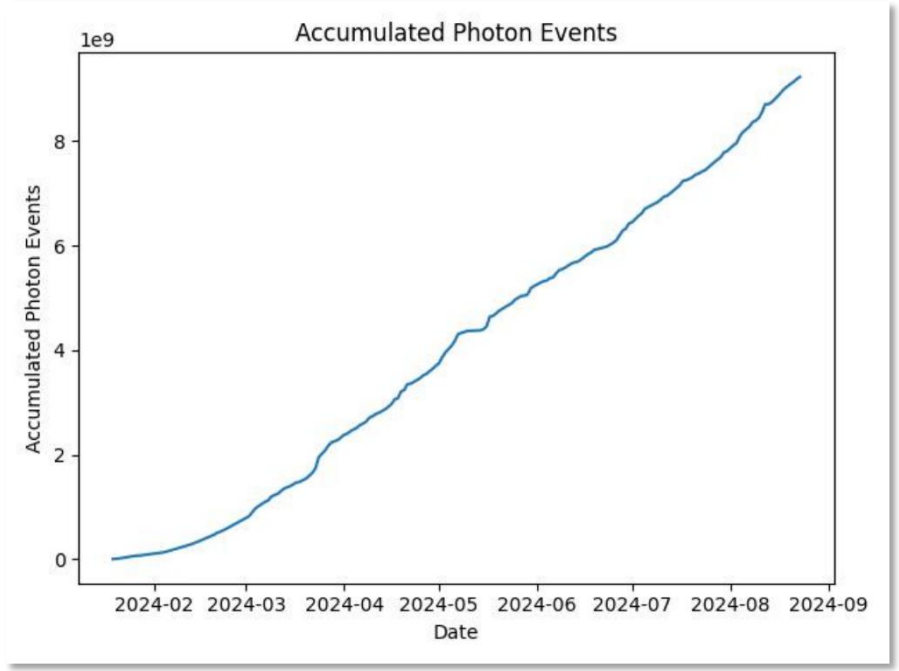
Time-bin之间的曝光不均匀性



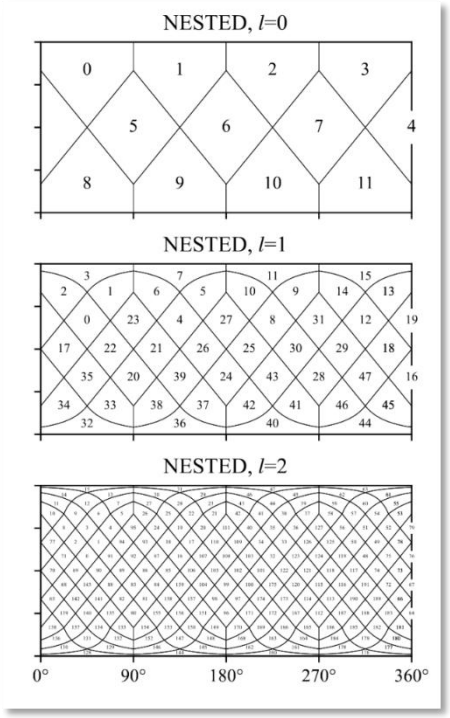
检索区域内的曝光不均匀性

数据量：300亿

National Astronomical Data Center
国家天文科学数据中心



50 million/day
20 billion/year



k	$N_{side} = 2^k$	N_{pix}	θ_{pix}
0	1	12	58°.6
1	2	48	29°.3
2	4	192	14°.7
3	8	768	7°.33
4	16	3072	3°.66
5	32	12,288	1°.83
6	64	49,152	55′.0
7	128	196,608	27′.5
8	256	786,432	13′.7
9	512	3,145,728	6′.87
10	1024	12,582,912	3′.44
11	2048	50,331,648	1′.72
12	4096	201,326,592	51″.5

Data Partition: HEALPiX Level 10
(pos_err of WXT is 5 arcmin)

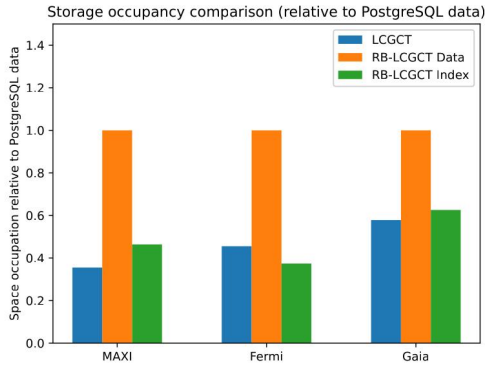
数据存储：TDengine集群

National Astronomical Data Center
国家天文科学数据中心

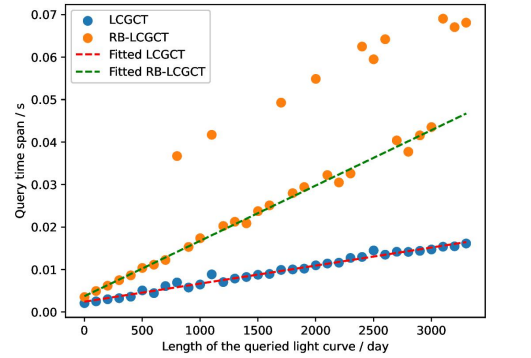
TimescaleDB：数据压缩能力不足

InfluxDB：高基数场景表现不佳

TDengine：已测试数据库中，最符合天文场景



存储效率

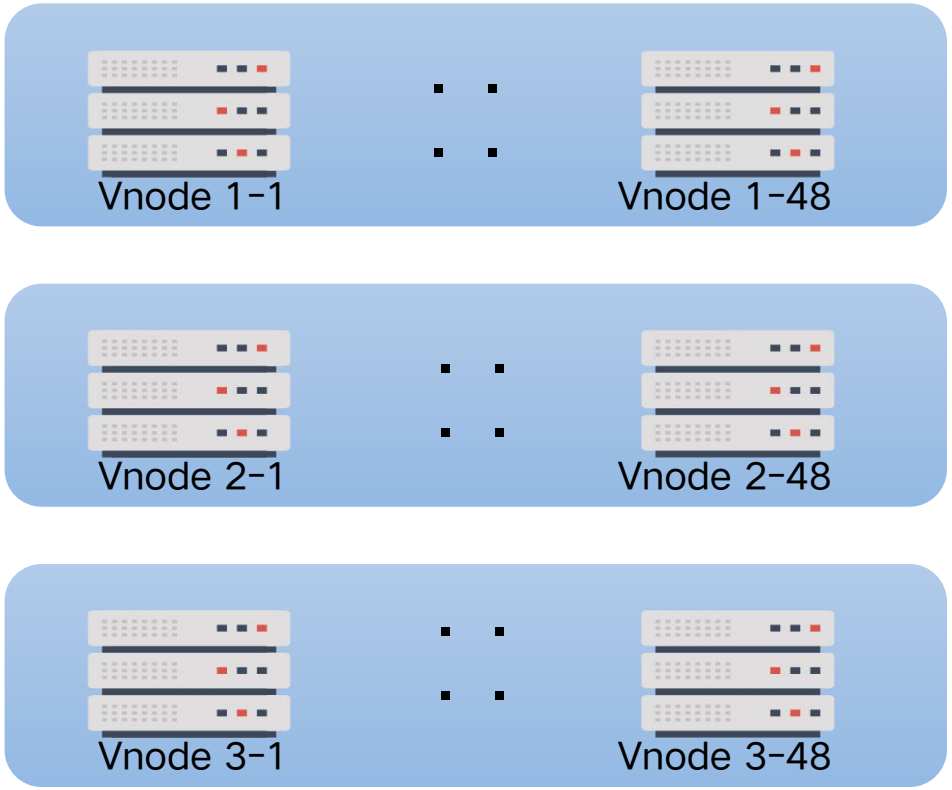


检索效率

数据存储：TDengine集群

National Astronomical Data Center
国家天文科学数据中心

time	x	y	ra	dec	healpix_id	arm_flag
2024-01-20 05:03:45.434	5514	6208	6.354897308972562	46.749765357985495	708280	false
2024-01-20 06:37:29.883	5510	6205	6.368372182435578	46.743148843364942	708280	false
2024-01-20 06:44:17.334	5509	6188	6.372845413014899	46.704587994103647	708280	false
2024-01-20 08:24:58.033	5504	6195	6.388949122741638	46.720729445675452	708280	false
2024-01-20 08:31:20.083	5509	6200	6.372029558792439	46.731842211316732	708280	false
2024-01-20 10:13:58.883	5522	6198	6.329054950090619	46.726656764974919	708280	false
2024-01-20 11:10:36.883	5515	6196	6.352405737585239	46.722462819977665	708280	false
2024-01-20 11:13:04.883	5513	6208	6.358214868022134	46.749814677909704	708280	false
2024-01-20 11:22:03.133	5512	6194	6.362490402293089	46.718068315323919	708280	false
2024-01-20 11:24:58.234	5511	6185	6.366420184405552	46.697676448028595	708280	false
2024-01-20 11:44:14.083	5516	6212	6.347986189122319	46.758750512458526	708280	false
2024-01-20 11:51:39.434	5523	6197	6.325808700085942	46.724335518672014	708280	false
2024-01-20 11:51:45.083	5519	6197	6.339072725976223	46.724535505458526	708280	false
2024-01-20 12:51:45.234	5512	6206	6.361669356538740	46.745321779178440	708280	false
2024-01-20 12:58:12.584	5510	6193	6.369189963938393	46.715895186931220	708280	false
2024-01-20 13:13:17.183	5497	6192	6.412360093546925	46.714251938353016	708280	false
2024-01-20 13:16:55.534	5506	6181	6.383263056045716	46.688835068740147	708280	false
2024-01-20 14:21:42.584	5519	6205	6.338517854779604	46.742704202868097	708280	false
2024-01-20 14:28:35.334	5507	6200	6.378662669038453	46.731939738441589	708280	false
2024-01-20 14:28:51.684	5506	6186	6.382925400134243	46.700191496947816	708280	false
2024-01-24 02:15:48.634	5918	6188	6.382435572688495	46.729500852542117	708280	false
2024-01-24 02:16:59.933	4215	3197	6.383674980513026	46.690636041672313	708280	false
2024-01-24 02:17:43.933	5911	6185	6.405983033909902	46.723306969881314	708280	false
2024-01-24 02:20:20.784	4219	3204	6.371303972386126	46.706713643543218	708280	false
2024-01-24 02:25:23.083	4219	3211	6.372108001237660	46.722495465538479	708280	false
2024-01-24 02:28:33.183	4214	3198	6.387084062780831	46.692816097408830	708280	false
2024-01-24 02:48:18.734	5923	6191	6.365507404233124	46.735866886507061	708280	false
2024-01-24 03:24:27.784	5933	6193	6.332139127660775	46.739515151370632	708280	false
2024-01-24 03:36:20.584	5917	6189	6.385624790886069	46.731857919937674	708280	false
2024-01-24 03:50:45.033	4233	3214	6.326304847238019	46.730286251029177	708280	false
2024-01-24 03:55:00.284	4221	3206	6.364943636846752	46.711370415601287	708280	false
2024-01-24 03:55:46.784	4220	3205	6.368123954555012	46.709042064284382	708280	false
2024-01-24 04:15:00.434	4226	3209	6.348809828162715	46.718501888322642	708280	false
2024-01-24 04:29:41.384	4208	3192	6.406148065349845	46.678842153284123	708280	false
2024-01-24 04:56:03.133	4228	3222	6.343695804406091	46.747959465095647	708280	false
2024-01-24 04:58:49.684	4230	3218	6.336647298563186	46.739086626990300	708280	false
2024-01-24 05:03:12.883	5931	6181	6.340255354227988	46.712465431284762	708280	false
2024-01-24 05:07:23.634	4213	3197	6.390262145692247	46.690487279598884	708280	false
2024-01-24 05:16:20.883	4225	3211	6.352333379078011	46.722937746744577	708280	false
2024-01-24 05:16:51.133	5930	6187	6.342820814982693	46.726168991638986	708280	false
2024-01-24 05:17:33.334	5914	6176	6.397146951649366	46.702621749160258	708280	false
2024-01-24 05:32:17.634	5922	6179	6.370296910074003	46.708725951670694	708280	false
2024-01-24 05:38:34.534	5921	6186	6.372746577271949	46.724698224364559	708280	false



Create table for each tile, Column Storage inside a table
(50,331,648 logical partitions)

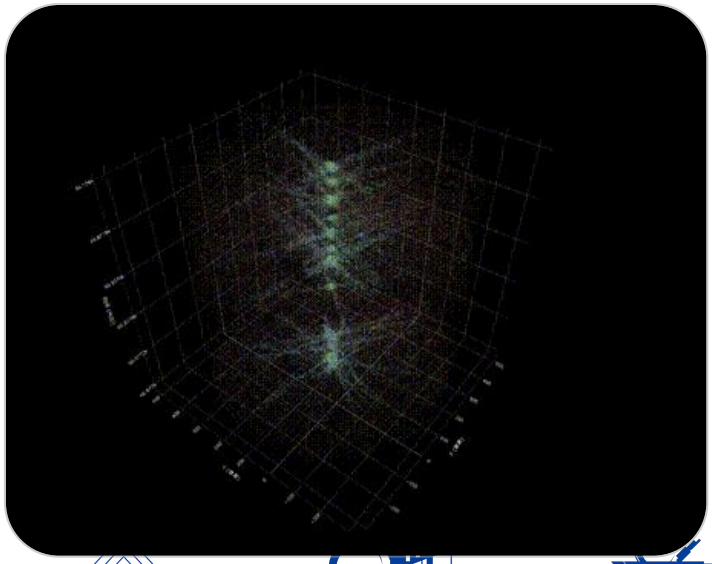
3 Server, 48*3 taos vnode
349,525 table each vnode

标准化的数据产品

National Astronomical Data Center
国家天文科学数据中心

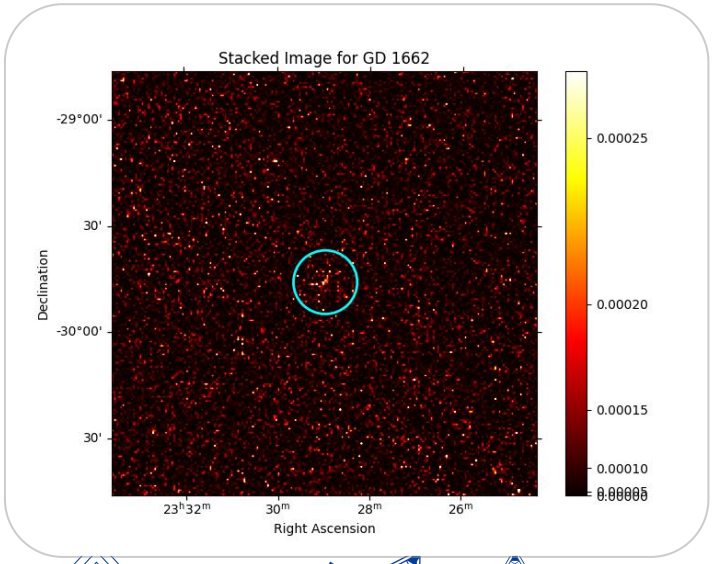
Cube

不同时间的图像组成的立方体



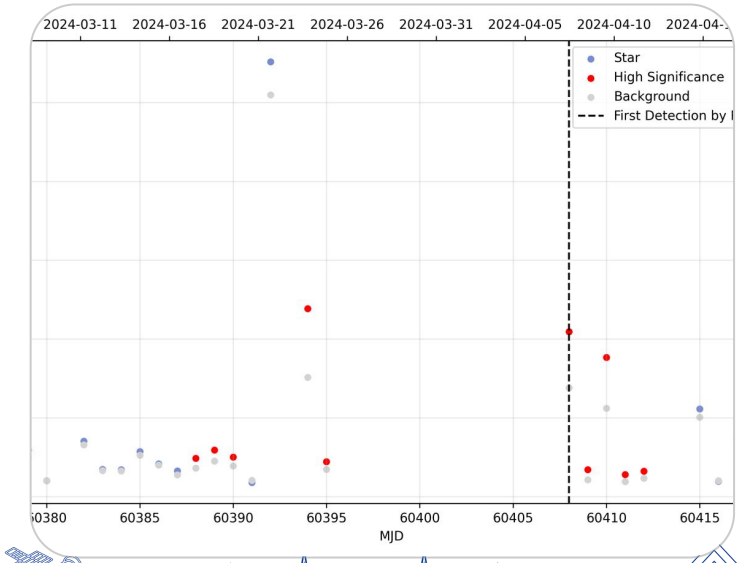
Image

长期叠加的图像



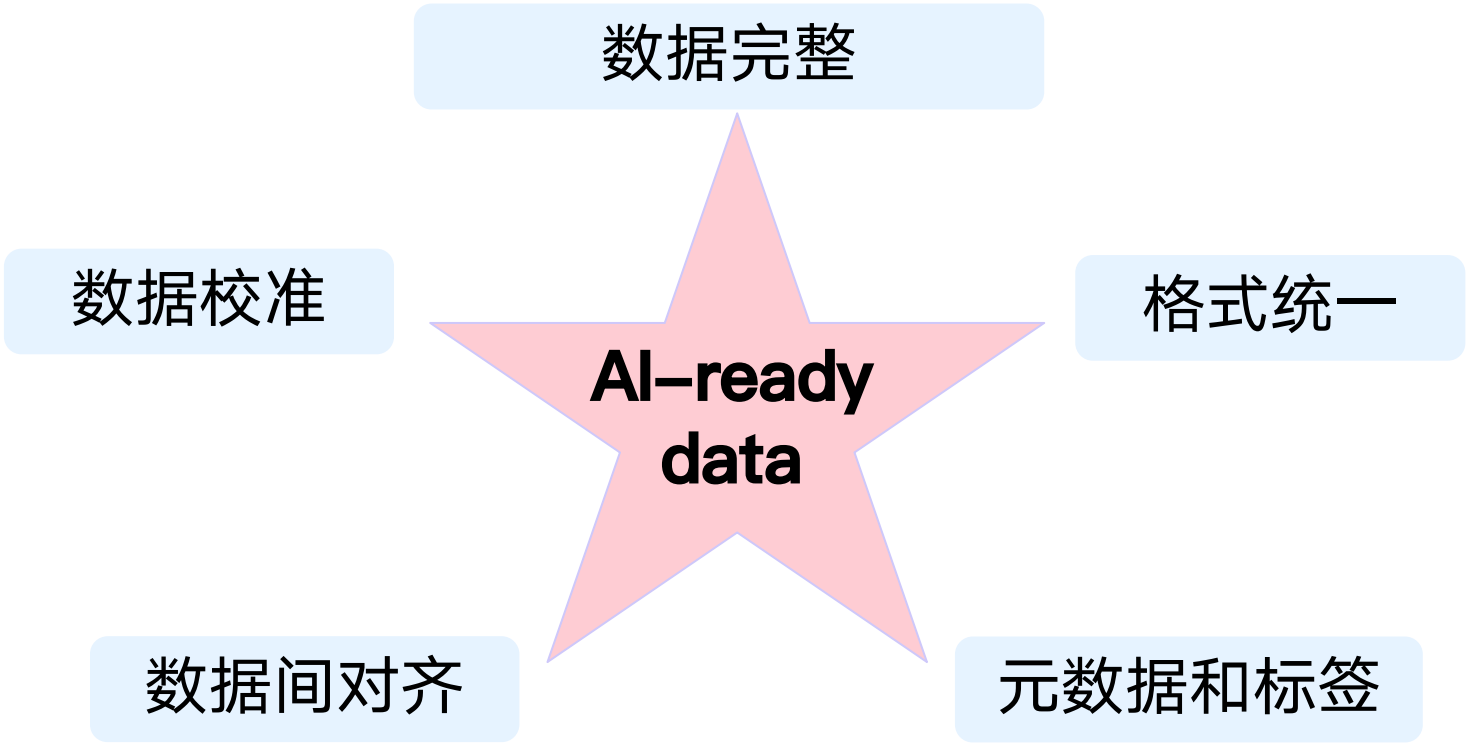
Light Curve

光子层级的光变曲线



光子数据库是AI-Ready的

National Astronomical Data Center
国家天文科学数据中心



Contents

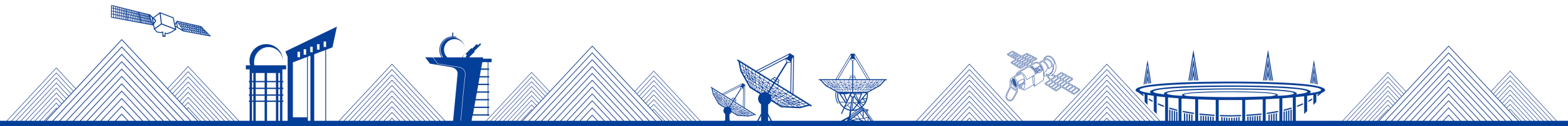
National Astronomical Data Center
国家天文科学数据中心

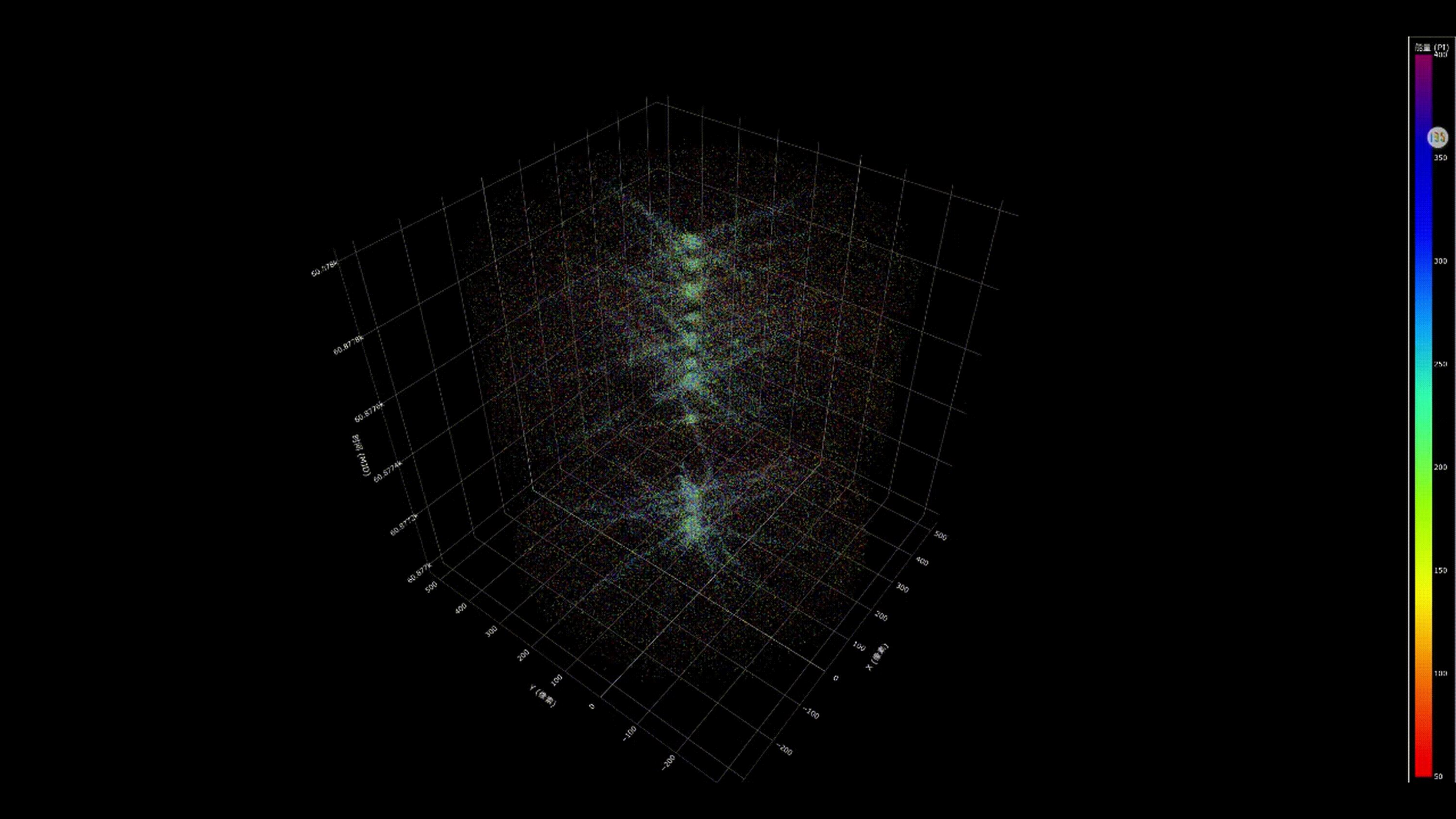
01 EP及其数据处理流水线

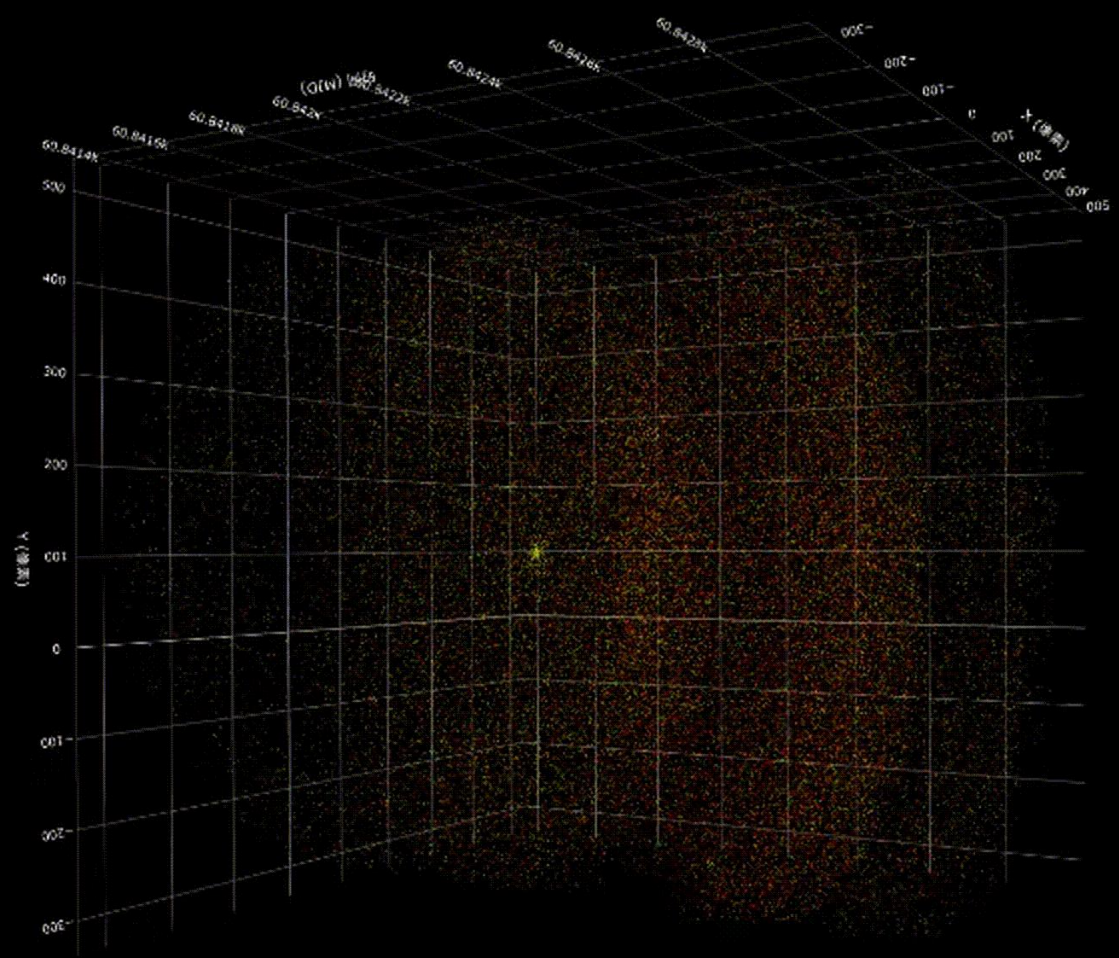
02 光子数据潜在价值

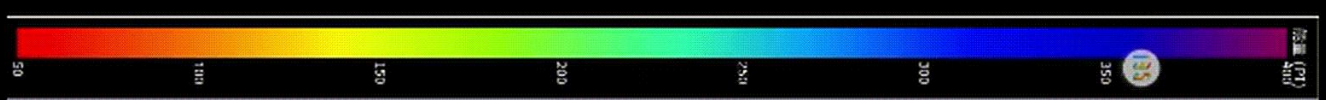
03 光子数据库的构建

04 应用场景与未来计划



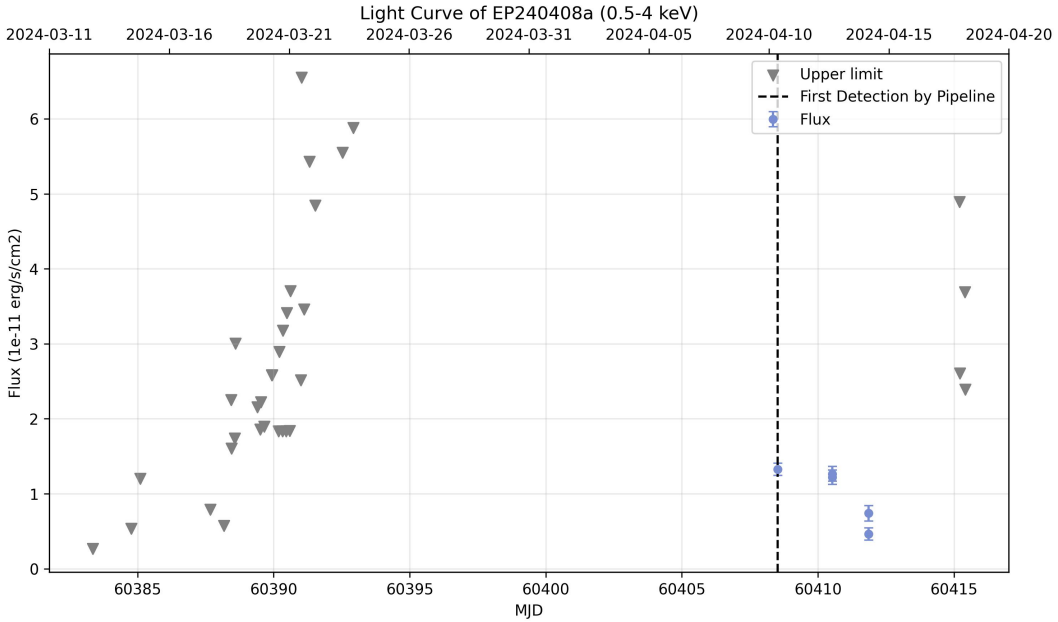

能量 (PI)
400
350
300
250
200
150
100
50
135
X (像素)
Y (像素)
500
400
300
200
100
0
-100
-200

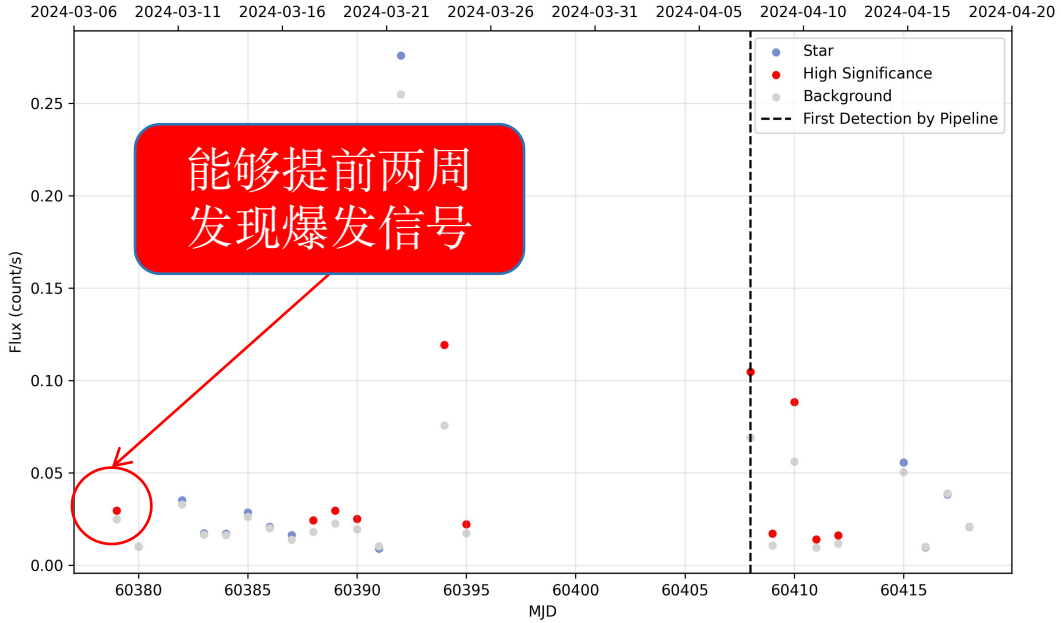

60.8424K
60.8422K
60.842K
60.8418K
60.8416K
(G(W)
-300
-200
-100
0
100
200
300
400
500
X (像素)
500
400
300
200
100
0
-100
-200
-300
Y (像素)


50
100
150
200
250
300
350
400

暂现源早期信号: 光变曲线

National Astronomical Data Center
国家天文科学数据中心



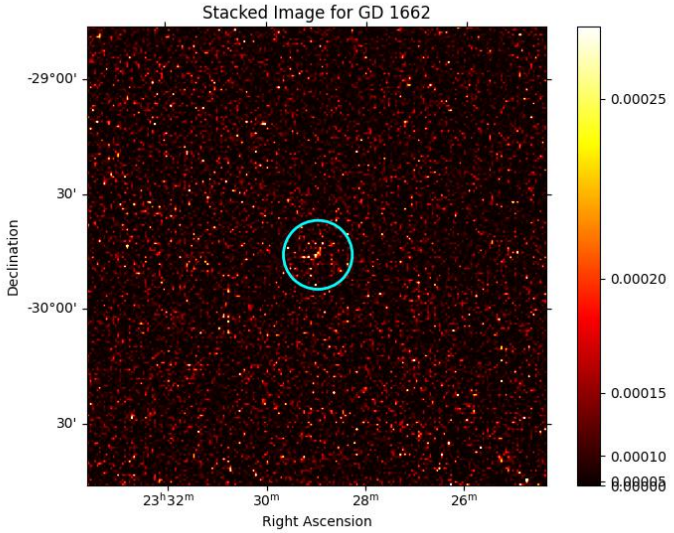


传统光变曲线（每个点是一次独立观测）

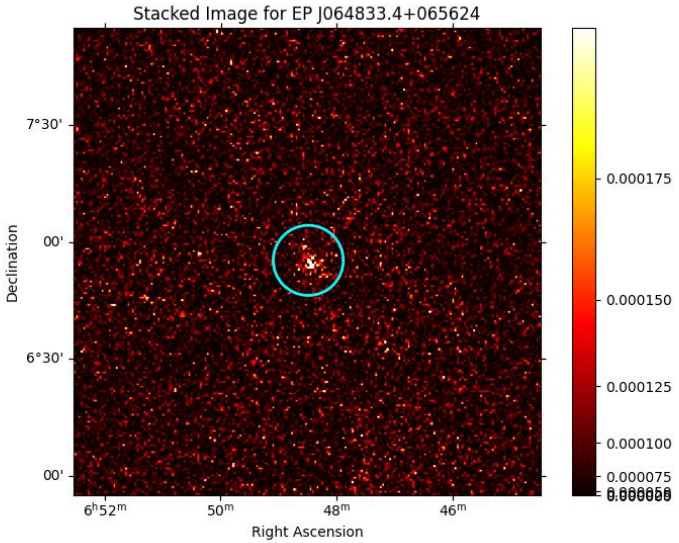
基于光子的光变曲线（每个点是一个time-bin）

暂现源早期信号: 叠加图像

National Astronomical Data Center
国家天文科学数据中心



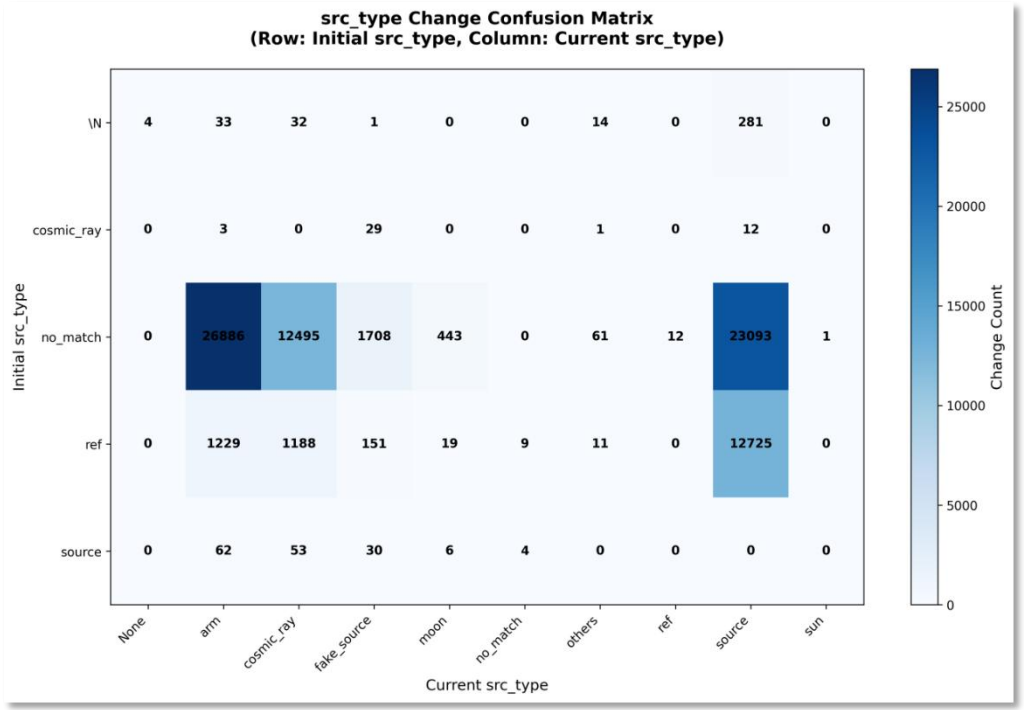
GD 1662



EP J064833.4+065624

两个暂现源在被流水线探测到之前的叠加图像

Arm（十字臂假源）识别



交叉匹配自动证认结果的混淆矩阵

问题

- 40%暂现源候选体最终被判定为Arm

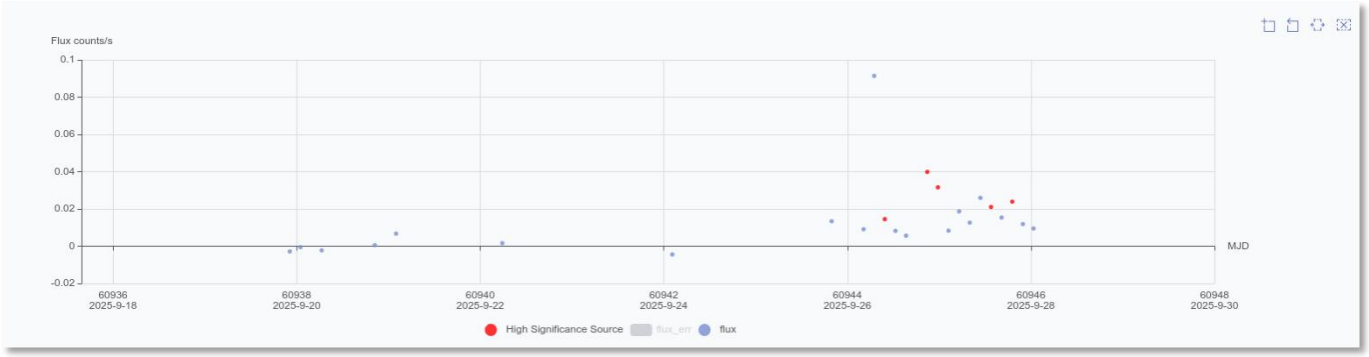
分析

- 局部的形态、能谱特征难以有效区分Arm和真源

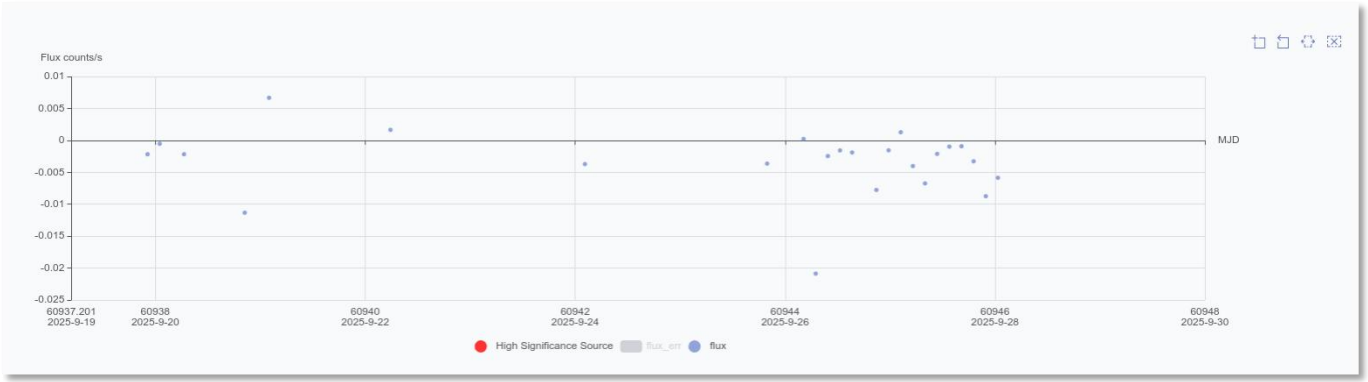
方法

- 依据PSF和星表构建图像，为每个光子标记arm_factor

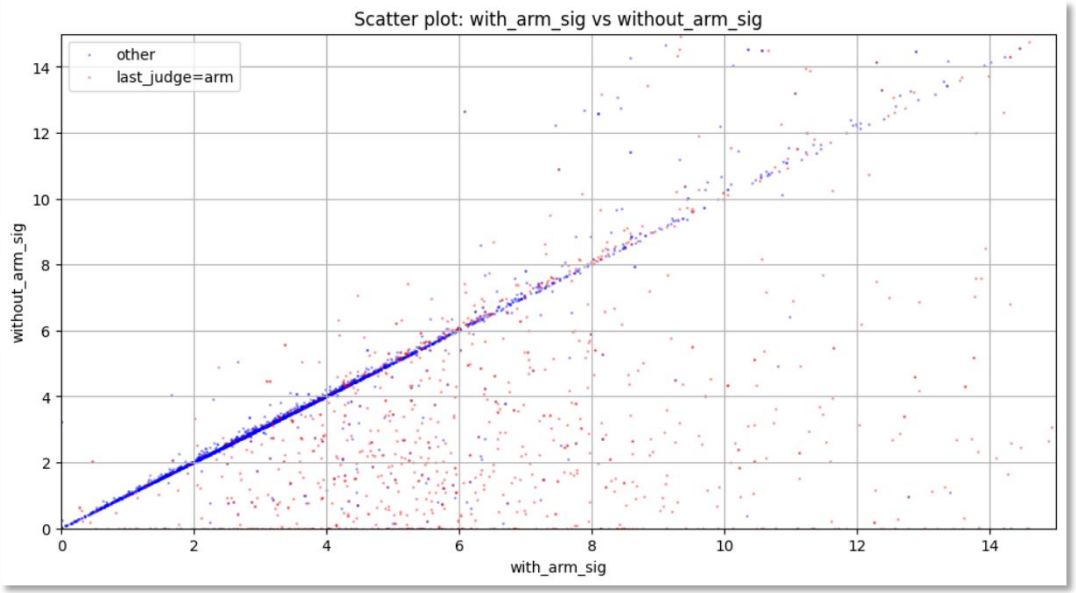
Arm识别



Binned light curve



Arm excluded light curve

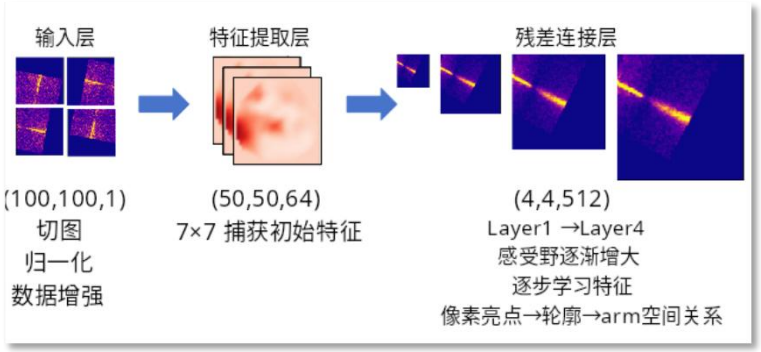


两种选项的置信度分布

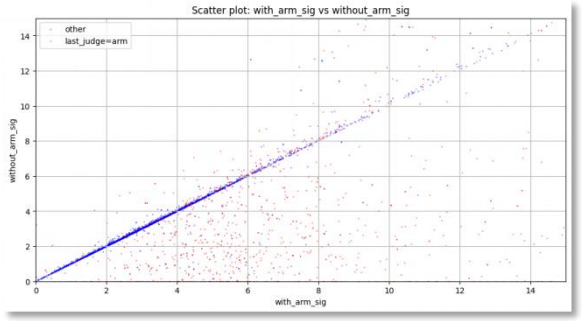
蓝色：真源

红色：十字臂

Arm识别：多通道CNN



基于CNN的形态识别算法（见陈朗海报）



基于arm_factor因子的识别

从光子数据库切出具有形态特征、“颜色”特征的多通道图形进行分类

计划：恒星耀发识别

National Astronomical Data Center
国家天文科学数据中心



恒星耀发

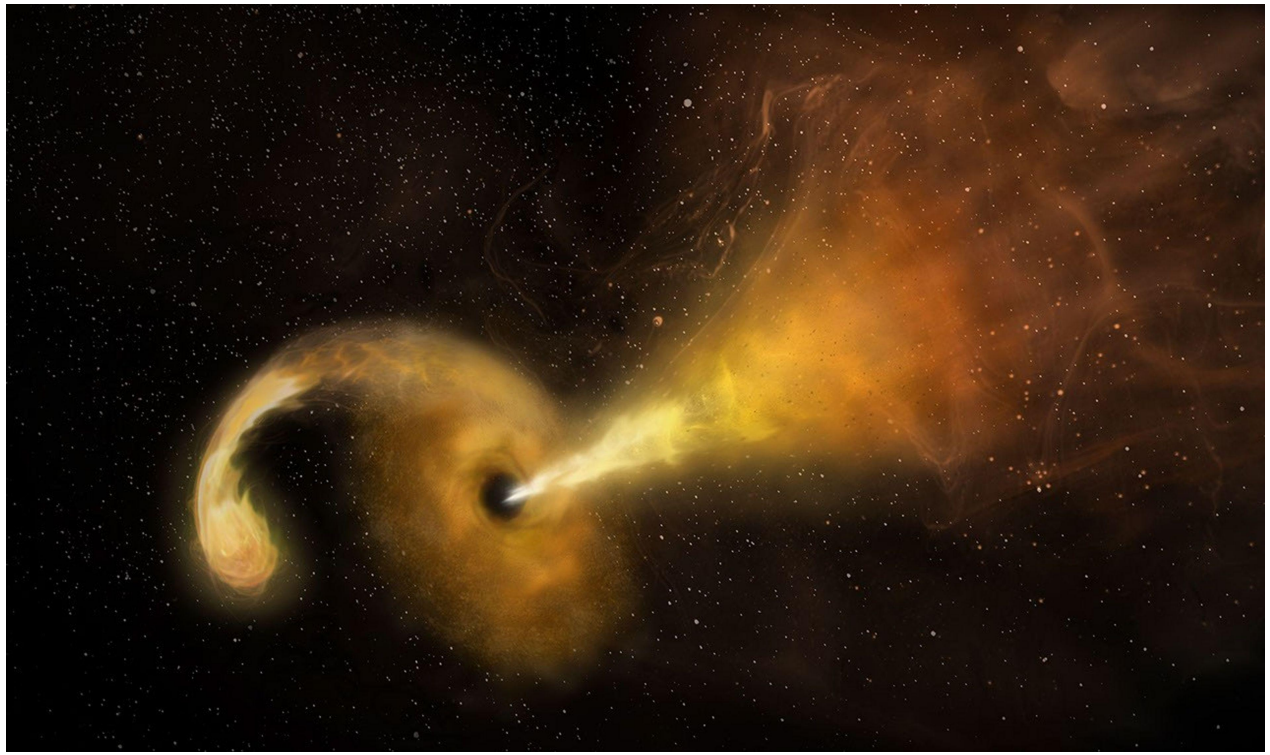
VS



伽马暴

短期特征相似，长期特征存在较大差异：例如单次或重复爆发、余晖特征等

计划：TDE搜寻



黑洞潮汐瓦解事件

问题

- 理论预期每年观测到10-130个TDE
- 2024年仅观测到2个TDE

分析

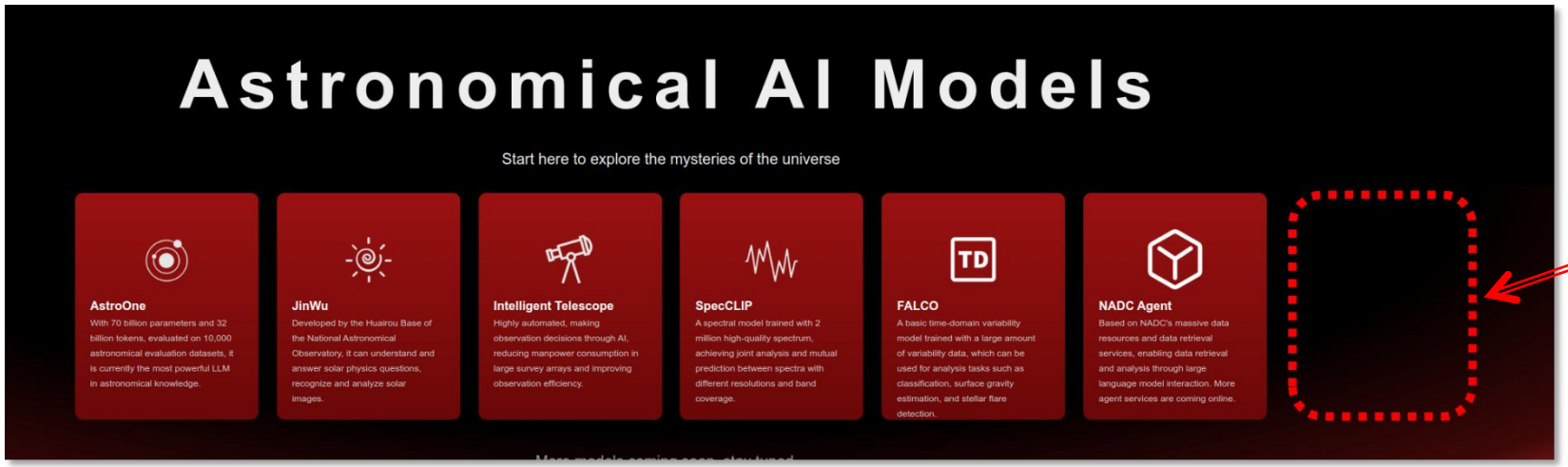
- 单次观测曝光过短
- 短期光变难以和AGN区分

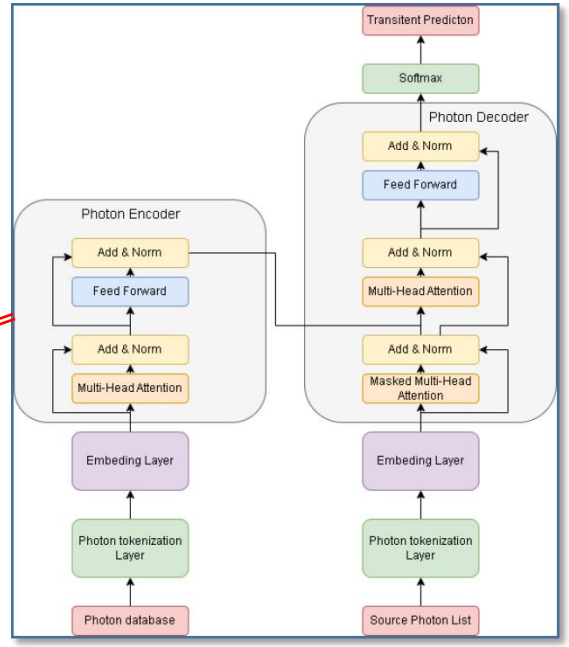
方法

- 光子级别的数据叠加和源探测
- 长期光变拟合

计划: 光子大模型

National Astronomical Data Center
国家天文科学数据中心





加入NADC 模型集，用于X射线源证认

基于Transformer的光子模型训练

Thank You!

