

doi: 10.3969/j.issn.1000-8349.2025.02.11

定标参考数据推荐方法

赵博伟^{1,2}, 卫守林³, 任娟娟¹, 王 锋⁴, 凌晨晓骥¹, 刘 超^{1,2}

(1. 中国科学院 国家天文台 空间天文与技术重点实验室, 北京 100101; 2. 中国科学院大学, 北京 100049;
3. 昆明理工大学 信息工程与自动化学院, 昆明 650500; 4. 广州大学 天体物理中心和国家科学数据中心大
湾区分中心, 广州 510006)

摘要: 自动化天文数据处理流水线的高效运行依赖于精准的定标参考数据推荐机制。因此, 完善的定标参考数据系统应运而生。系统地回顾了国际主流望远镜采用的定标参考数据推荐方法, 并深入分析了各方法的优缺点。重点介绍了一种基于文本规则的新型推荐策略和配套的定标参考数据系统, 以及其在自动化数据处理中的灵活性和高效性。此外, 详细阐述了该系统在中国空间站工程巡天空间望远镜 (CSST) 科学数据处理中的关键作用和潜在应用价值, 并展望了其未来的发展前景。该研究为天文数据处理中定标参考数据的推荐提供了新的思路和方法, 具有重要的理论和实践意义。

关 键 词: 天文数据处理; 定标参考数据; 空间天文望远镜

中图分类号: P171

文献标识码: C

1 引 言

天文学是研究宇宙中各类天体及天文现象的学科^[1], 其研究高度依赖观测^[2]。通过观测, 可获取丰富的数据, 这些数据经过必要的处理后, 才能用于测量和分析, 以揭示天体的基本物理特性, 如亮度、温度、光谱型、视向速度、化学成分及结构等。这些关键数据不仅为深入探索宇宙及各类天体的形成与演化提供了重要支撑, 也奠定了现代天文学中天体物理学的核心地位^[3]。天文观测技术经历了从人眼观测与手绘天图到现代天文望远镜的跨越式发展, 逐步形成了包括天体照相术、天体测光术、天体分光学、全波段天文学以及数字化巡天在内的全方位观测体系^[4]。与此同时, 天文望远镜也实现了从小口径到大口径、从单镜面到多镜面、从地基到空间的巨大飞跃^[4, 5]。如今, 现代天文观测主要依托大口径天文望远镜及其配套的先进终端仪器, 为天文学研究提供了前所未有的探测能力^[6]。

收稿日期: 2024-04-15; 修回日期: 2024-07-12

资助项目: 中国载人航天工程巡天空间望远镜专项科学研究 (CMS-CSST-2025-A19); 国家自然科学基金 (12090041, 12090040)

通讯作者: 刘超, liuchao@nao.cas.cn

天文望远镜的应用不仅仅是收集观测数据, 其核心目标在于推动科学研究, 进而创造新知^[7]。因此, 将原始观测数据转化为科学知识是至关重要的过程, 数据处理与分析软件则在其中发挥着桥梁作用^[7]。通常, 天文观测的原始数据不仅包含来自天体的真实信号, 还包括了大气和仪器效应的影响^[8], 并混杂了各种来源的噪声^[9]。因此, 最初获取的原始观测数据需要经过仪器效应改正、宇宙线剔除、各类定标等一系列复杂的数据处理过程, 才能转化为可供天文学家分析研究的科学数据。然而, 随着现代天文观测数据处理的复杂性不断增加, 单个终端仪器往往会涉及多种类型的数据, 包括科学观测数据和定标数据。例如, 欧洲南方天文台 (European Southern Observatory, ESO) 甚大望远镜 (Very Large Telescope, VLT) 的第二代终端仪器——宽带中分辨率光谱仪^[10] (X-shooter), 其原始观测数据的处理涉及上百种数据类型^[9, 10]。此外, 不同的天文望远镜在设计建造之初均围绕特定的科学目标展开。为达到这些目标并为科学家提供高质量的科学数据, 对终端设备和科学观测数据进行精确的定标是最基本且至关重要的环节。

完整的定标流程包括定标观测、定标数据处理 (即生成定标参考文件) 以及如何选择最佳的定标参考文件进行科学数据处理。由于获得理想的定标观测数据通常需要消耗大量观测时间, 而且观测时间的分配策略会直接影响后续科学数据处理中定标参考数据的选择规则。因此, 究竟是开展更多的定标观测, 还是投入更多时间进行科学观测, 不可避免地需要进行权衡。对于地基天文观测, 理想的情况下, 应使用当晚观测的定标数据对当晚的科学观测数据进行定标。然而, 若终端仪器具有较高的稳定性, 则前几晚或后几晚的定标数据在一定程度上也可用于当晚科学观测数据的定标, 以提高观测效率。

空间天文望远镜位于地球大气层之外, 能够在太空环境中开展全波段、高精度、长时间的观测, 从而获得高质量的数据^[11-13]。然而, 复杂的空间环境^[14]可能导致探测器受到辐射损伤, 引发仪器性能改变^[15], 进而影响数据处理和科学产品的质量。因此, 空间望远镜的在轨定标至关重要^[16]。国际知名空间望远镜, 如哈勃空间望远镜 (Hubble Space Telescope, HST)^[17-19]、韦布空间望远镜 (James Webb Space Telescope, JWST)^[20-22]、罗曼空间望远镜 (Nancy Grace Roman Space Telescope)^[24, 25]等, 均设有周期性的在轨定标观测。此外, 天文望远镜通常配备多种终端仪器, 每种终端都具有独特的观测模式, 并需要特定类型的定标观测^[10, 20]。不同类型的定标观测因其目标不同, 通常具有不同的监测周期, 这进一步增加了定标参考数据的复杂性。

随着天文观测数据量的激增, 构建自动化科学数据处理流水线已成为必然趋势^[9, 23]。为了确保科学数据的准确处理, 自动化流水线的设计不仅需要能够自动识别并选用最佳的定标参考文件^[28], 还需综合考虑科学观测策略与定标方案等因素。因此, 在科学数据处理过程中, 如何高效、准确地选择最佳的定标参考数据, 成为既复杂又关键的问题。针对这一需求, 多个国际望远镜项目开发了专门用于推荐最佳定标参考数据的系统, 尤其是在更为复杂的空间望远镜任务中, 这类系统发挥着至关重要的作用^[28-31]。其中, JWST 的定标参考数据系统 (Calibration Reference Data System, CRDS) 因其独特性和先进性而尤为突出。基于 HST 长期积累的经验, 为满足 JWST 数据处理需求, 一个以构建定标参考文件推荐规则和自动化推荐最佳定标参考文件功能为核心, 同时兼顾定标参考文件存储和管理功能的定

标参考数据系统应运而生^[28]。该系统作为连接定标参考文件与科学数据处理流水线的核心纽带,能够按照既定规则,从多类型、多版本的大量在轨定标参考文件中自动推荐出最佳选项,供科学数据处理流水线使用。此外,基于文本规则的定标参考数据系统具有极高的灵活性和可推广性,其不仅适用于天文望远镜观测数据的定标参考文件选择,还可应用于其他需要基于规则进行自动化推荐的场景^[28]。

本文第2章将系统回顾和总结国际上各望远镜主要采用的定标参考数据推荐方法;第3章详细介绍当前较为灵活和先进的基于版本化文本规则的定标参考数据系统;第4章重点阐述中国空间站工程巡天空间望远镜(China Space Station Survey Telescope, CSST)对定标参考数据系统的应用;第5章对全文进行总结与展望。

2 历史回顾

基于科学数据处理的需求,在实际天文观测过程中,针对不同的终端仪器和观测模式,望远镜会周期性地穿插进行各类必需的定标观测。这些周期性拍摄的各类定标观测数据,经过处理会生成不同版本的各类定标参考文件。因此,在后续进行大批量天文观测数据自动化处理的过程中,如何从大量不同版本的各类定标参考文件中,自动选用最佳的文件进行科学观测数据的定标处理,即如何为科学数据处理流水线自动分配最佳定标参考文件,会显著影响数据处理结果和精度,这需要重点研究。

关于定标参考数据的推荐方法,各望远镜项目采用了多样化的解决方案。通常比较简单的方法是,通过结构化查询语言(Structured Query Language, SQL)从数据库^[32]中检索出与科学数据观测时间最为接近的定标参考文件^[29, 30],以供科学数据处理流水线使用。但该方法存在诸多弊端,如历史处理结果难以再现,过于依赖数据库,无法传播使用,难以根据特殊需求定制修改所用的参考文件等^[28]。因此,国际上各天文望远镜根据各自实际需求,采用了不同的定标参考数据推荐方法。

钱德拉 X 射线天文台(Chandra X-ray Observatory, CXO)^①是美国航空航天局(National Aeronautics and Space Administration, NASA)于1999年发射的一颗 X 射线天文卫星,也是大型轨道天文台计划的第三颗卫星。CXO 使用的是一种基于“定标数据库”(chandra calibration database, CALDB)^②的方法^[29],该方法是根据 NASA 高能天体物理科学档案研究中心(High Energy Astrophysics Science Archive Research Center, HEASARC)开发的 CALDB 系统标准模型构建的,大部分高能天文望远镜使用的都是该方法。尽管 CALDB 方法的名称中似乎包含数据库,但其本质上并未涉及到真正的数据库。CALDB 方法的核心在于选择定标参考文件的规则保存在索引文件(index files)里,索引文件列出了所有可用的定标参考文件和使用条件。虽然该方法构建起来并不简单,但也有其优势,如可再现过去的推荐及软件可传播等。

^①<https://chandra.harvard.edu/>

^②<https://cxc.cfa.harvard.edu/caldb/>

斯皮策空间望远镜 (Spitzer Space Telescope, SST)^①是 NASA 于 2003 年发射的望远镜, 是大型轨道天文台计划的最后一台空间望远镜, 其灵敏的红外探测能力使其能够观测到宇宙中一些最遥远和最暗淡的天体, 揭示了宇宙的许多秘密^[33, 34]。在进行科学数据处理时, SST 使用 “caltrans” (calibration transfer 的缩写) 软件进行所需定标参考文件的检索和选择, 该软件由斯皮策科学中心 (Spitzer Science Center, SSC) 开发, 使用 C 语言实现, 并与 Informix 数据库交互。SST 使用的可能是最为纯粹的基于数据库的方法^[30], 其定标参考文件的选择规则基本上表现为对数据库的 SQL 查询。虽然该系统支持简单的更新、撤销操作, 但其难以再现以往的推荐结果 (参考文件), 不易于传播使用, 无法在其他环境中运行。

双子座望远镜 (Gemini Telescope)^②是由美国大学天文联盟 (Association of Universities for Research in Astronomy, AURA)^③负责实施的两台 8 m 级地基光学天文望远镜, 于 2000 年投入使用。其中双子南座望远镜位于南半球智利安第斯山脉的帕穹山, 双子北座望远镜位于北半球夏威夷莫纳克亚山。与 CXO 和 SST 相比, 双子座望远镜采用的是一种混合方法: 多数信息使用数据库存储, 但定标参考文件的选择则是通过 Python 代码生成的 SQL 查询语句来对数据库进行检索^[35]。简单地使用 SQL 查询通常无法满足一些选择规则比较复杂的情况下的需求, 而这种混合方法则可构建出任意复杂的规则。虽然它也提供了一个网页服务^④来供外部用户选择定标参考文件, 但该系统不允许传播, 不支持多版本的规则, 用户不易修改定制。

由美国和欧洲联合研制的 HST^⑤以美国天文学家 Edwin Hubble 命名, 是一架大型空间光学望远镜, 于 1990 年搭载发射号航天飞机发射升空, 并运行于近地轨道。为了更好地支撑科学数据处理和分析, 特别是科学数据定标流水线的运行, HST 在发射前就开始规划和设计定标数据库系统 (calibration database system, CDBS)^[31, 36-38], 该系统成功地被 HST 使用了 20 余年, 在此期间, 虽然其工作方式进行了完善和升级, 但整体设计框架基本保持不变^[39]。

CDBS 的基本设计是以数据库为中心的, 数据库包含了所有类型定标参考文件的信息^[39]。对于 HST 所有终端设备的每个定标参考文件, CDBS 中都会添加一个文件描述。文件描述中含有定标参考文件所使用的能够定义曝光类型的终端设备观测模式参数, 这些描述信息都保存在每个 HST 终端设备的单独数据库表格中, 每个表包含了该终端设备的定标参考文件集的所有观测模式的可选择参数。定标参考文件描述信息还包含 “UseAfter” 时间戳, 该时间戳标记了该参考文件应该被使用的最早时间^[38, 39]。此外, 所有最佳定标参考文件的选择规则都存储在 XML (eXtensible Markup Language, 可扩展标记语言) 文件中,

^①<https://www.spitzer.caltech.edu/>

^②<https://www.gemini.edu/>

^③<https://www.aura-astronomy.org/>

^④<https://archive.gemini.edu/searchform>

^⑤<https://www.stsci.edu/hst>

这些规则用来指定为科学数据推荐参考文件时所使用的搜索条件。

CDBS 的主要功能是, 为每个正在处理的新科学数据集 (或根据情况需要重新处理的科学数据集) 在数据库中查询所需的各种类型的定标参考文件, 并返回一个最佳参考文件的文件名, 然后使用 “bestref” 程序将这个名称更新为科学数据集头文件中特定关键词的值^[39]。科学数据处理流水线从科学数据集头文件关键词中检索出需要的最佳参考文件, 用于科学数据的定标。

虽然 CDBS 被成功地应用于 HST 20 余年, 但在长期的使用过程中, 人们发现其存在着诸多局限^[28]。例如, 难以对新的参考文件进行测试检查, 提交的参考文件若存在错误也难以撤销, 不支持远程使用, 不支持用户对参考文件选择规则的个性化定制需求, 很难或几乎不可能再现历史结果等。由于选择规则存储在数据库里, 真正有效的选择规则难以被理解。以数据库为核心的设计理念, 也限制了选择规则的类型扩展, 使得增加新类型的选择规则难以实现。这些局限归根结底在于 CDBS 对数据库的依赖, 即将选择规则存储于数据库^[39]。

鉴于 CDBS 潜在的诸多局限, 美国空间望远镜科学研究所 (Space Telescope Science Institute, STScI)^①为 JWST^②开发了一套新的系统, 即定标参考数据系统 (Calibration Reference Data System, CRDS)^[28], 该系统使用文本来囊括参考文件的选择规则, 有效地解决了对数据库的依赖问题, 也实现了系统的可移植性。该系统在 JWST 上试验成功后, HST 也逐步由 CDBS 迭代更新为了 CRDS^[28]。NASA 下一代旗舰级空间任务罗曼空间望远镜^③, 也将使用该定标参考数据系统。与通用型空间天文望远镜 JWST 和 HST 不同, 罗曼空间望远镜兼具巡天功能^[40]。表 1 总结了国际上几台望远镜使用的定标参考数据推荐方法。本文将在第 3 章对全新的定标参考数据推荐系统 CRDS 进行详细介绍。

表 1 国际上各望远镜所用的定标参考数据推荐方法

望远镜名称	运行时间	定标参考数据推荐方法
钱德拉 X 射线望远镜	1999 年 7 月至今	定标数据库 (CALDB) ^[29]
斯皮策空间望远镜	2003 年 8 月—2020 年 1 月	定标转换 (caltrans) ^[30]
双子座望远镜	2000 年至今	数据库+Python 查询的混合方法 ^[35]
哈勃空间望远镜	1990 年 4 月至今	前 24 年定标数据库系统 (CDBS) ^[37, 39] , 现在定标参考数据系统 (CRDS) ^[28]
韦布空间望远镜	2021 年 12 月发射, 计划运行 10 年	定标参考数据系统 (CRDS) ^[28]
罗曼空间望远镜	预计 2026 或 2027 年发射, 运行 5 年	定标参考数据系统 (CRDS) ^[28]

^①<https://www.stsci.edu/>
^②<https://webbtelescope.org/>
^③<https://roman.gsfc.nasa.gov/>

3 定标参考数据系统

3.1 概述

随着空间天文的发展,为了更好地服务自动化天文数据处理,科学家设计了适用于 JWST、HST、Roman 三大空间望远镜的定标参考数据系统 (CRDS),它是一套新的用于管理和分配定标参考文件的系统。对于科学数据处理流水线,CRDS 管理的参考文件包括两类,一类是包括定标数据或信息的“数据参考文件”(data reference files),另一类是科学数据处理流水线中用到的包含各类配置参数信息的“参数参考文件”(parameter reference files)^[41, 42]。图 1 展示了罗曼空间望远镜终端 WFI (Wide Field Instrument) 的二级数据处理流水线(用于生成定标后的单次曝光数据)及其与 CRDS 之间的关系。CRDS 之所以被称为“系统”,是因为围绕其核心功能即“最佳参考文件推荐功能”,需要构建完备的服务支撑平台,最终使它成为了一个包含 Python 库、命令程序、网络服务器和数据库的综合“系统”,用于管理和分配某个天文望远镜所有终端仪器的所有类型的定标参考文件,高效灵活地服务于自动化科学数据处理。

CRDS 经过全新设计,摒弃了使用数据库存储定标参考文件选择规则的传统解决办法^[39],而是采用更简洁且灵活的方法,即将一个版本的所有选择规则都囊括在一个简单的文本文件里^[28]。该方法有着多项优势:首先,文本化形式使得各类型定标参考文件的选择规则清晰明了,易于理解;其次,这些规则文件是版本化控制的,不同版本之间互不冲突,无混淆,易于管理,历史处理结果可再现;再次,文本格式的规则文件便于传播,因此远程用户无需安装复杂软件即可轻松使用;最后,规则文件的内容较为简单,必要时,用户还可根据自身实际需求方便地定制和修改规则,灵活易操作。

为了实现 CRDS 的核心功能“最佳参考文件推荐”,前提需构建各类定标参考数据的文本化规则文件,即基于这些规则文件来实现最佳参考文件的推荐。作为一套完备的系统,为支撑该需求,CRDS 需要具备多方面的功能,包括新版参考文件的提交并同时实现新版规则文件的重构、最佳参考文件推荐、科学数据再处理、参考文件和规则文件的检查验证等其他各类辅助功能(详见表 2)。图 2 为 CRDS 各项功能的整体示意图。可以看出,CRDS 不同的功能涉及的用户不同。CRDS 涉及到的用户大体有三类:定标科学家、科学数据处理流水线运行科学家、CRDS 系统管理者。其中,定标科学家主要是制作和生成各类定标产品即定标参考文件,并提交到 CRDS,为 CRDS 提供输入;科学数据处理流水线运行科学家,则是 CRDS 最佳参考文件推荐功能的使用者;CRDS 系统管理者则负责整个 CRDS 系统的维护和管理。

除了一些网络交互工具,CRDS 主要是基于 Python 开发的^[28]。围绕以上功能,实现了脚本命令行工具、Python API (Application Programming Interface) 接口及简单的可视化

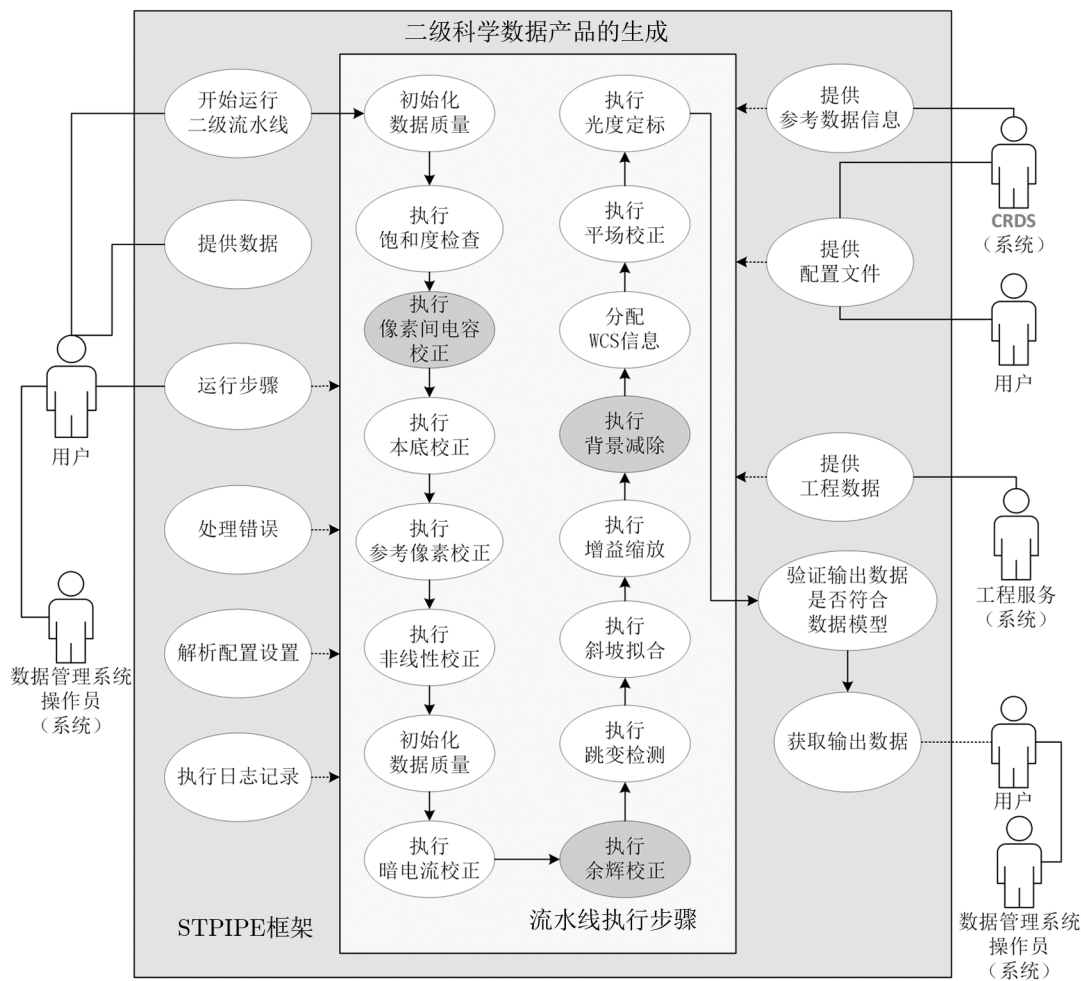


图 1 罗曼空间望远镜终端 WFI 二级科学数据产品生成流程及与 CRDS 的关系^[42]

网页界面^{①②③}等多种使用方式。CRDS 软件有客户端版本和服务端版本，github 上释放的为其客户端版本^④，其服务端软件（主要控制文件提交、管理、网络服务等）因安全方面的考虑，未公开释放。通过 CRDS 客户端软件或者网站，用户可以使用其参考文件提交与检查、最佳参考文件推荐、科学数据再处理及各类辅助功能（见表 2）。

^①<https://jwst-crds.stsci.edu/>

^②<https://hst-crds.stsci.edu/>

^③<https://roman-crds.stsci.edu/>

^④<https://github.com/spacetelescope/crds>

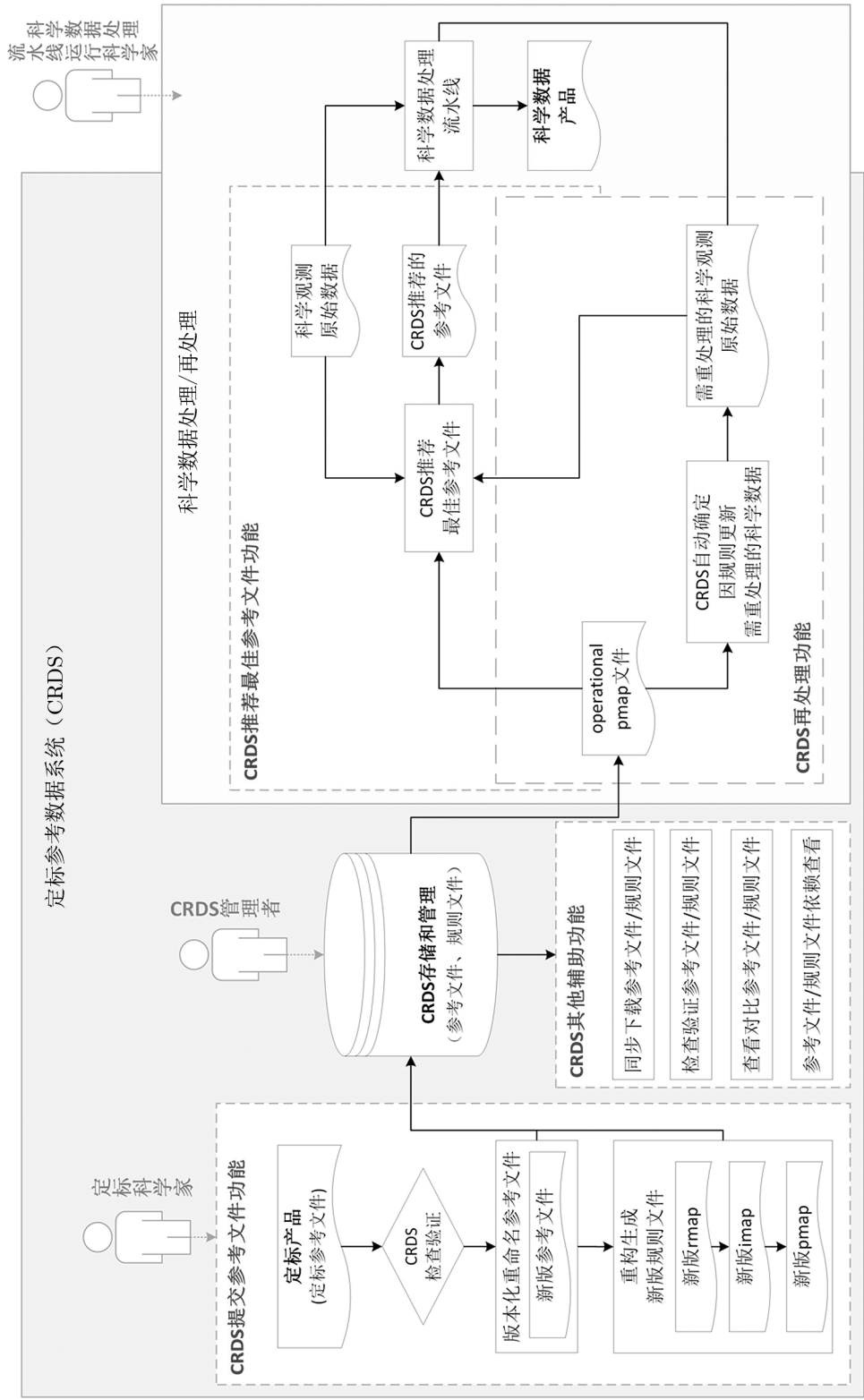


图 2 CRDS 功能汇总示意图

表 2 CRDS 各类功能详细介绍

功能	详细描述
推荐最佳参考文件 (bestrefs)	为科学数据集推荐出最佳定标参考文件，有文件模式、再处理模式、规则文件测试模式、比较模式
验证检查 (certify)	用于参考文件/规则文件的检查验证，包括基本文档格式、语义、参数限制等检查
校验 (cheksum)	用于添加、删除或验证规则文件的哈希值 (sha1sum)，或参考文件的校验和 (checksum)、数据和 (datasum)
比较 (diff/rowdiff)	对不同版本的规则文件或参考文件进行对比，并返回文件之间的差异；diff 对 fits 文件中的表格数据逐列比较，rowdiff 则专门用来逐行比较表格数据
下载参考文件 (get_synphot)	从 CRDS 服务器中下载某版 pmap 文件能够映射的参考文件到本地
再处理 (query_affected)	自动确定因规则或参考文件更新需重新处理的科学数据集
信息查询 (list)	用于列出或查询各类 CRDS 信息，如本地安装 CRDS 的配置信息、服务器端或本地缓存中的参考文件/规则文件等
匹配 (matches)	用于解释 CRDS 规则，找出对某个参考文件匹配的规则内容
生成context文件 (newcontext)	自动生成新版 context 文件 (imap 文件或 pmap 文件)
重构 (refactor)	自动重构新版的 rmap 文件
提交 (submit/rc_submit)	提交参考文件到 CRDS 服务器端（此功能需要有 CRDS 账号权限）
同步 (sync)	将服务器端的参考文件/规则文件同步到本地
重命名 (uniqname)	用于重命名具有 HST 唯一官方 CRDS 名称的参考文件
依赖查看 (uses)	用于查看、列出某参考文件/规则文件的依赖文件

3.2 选择规则文件

CRDS 系统围绕选择规则文件展开，选择规则文件描述了如何为科学数据选择或分配最合适的定标参考文件。CRDS 的规则文件是层级化设计的文本文件，规则文件之间的层级映射关系通过逐级映射，帮助科学观测数据简单清晰地选出所需的最佳定标参考文件。其规则层次结构有三层（见表 3），即 pmap 文件（对应望远镜流水线层级，一台望远镜对应一个 pmap）、imap 文件（对应望远镜所装载的终端仪器层级，每个终端仪器对应一个 imap）、rmap 文件（对应终端仪器的定标数据类型层级，每类定标参考文件对应一个 rmap），这三种规则文件具有层级映射/选择关系，因此都被称为映射（mapping）文件，其中 imap 文件和 pmap 文件又被称为上下文（context）文件。基于这些规则文件，CRDS 实现与不同级

别的功能相对应的嵌套组织结构。图 3 给出了 CRDS 规则文件的层级结构。望远镜某一终端仪器某种定标类型的所有观测模式的参考文件囊括在版本化的 rmap 文件中 (见图 3 中的第 3 和 4 列); 望远镜某终端仪器的 imap 文件收集该终端仪器所有定标类型的某版本的 rmap 文件, 这些文件构成了 imap 文件的有效推荐版本 (见图 3 中的第 2 和 3 列); 望远镜流水线层级的 pmap 文件则收集该望远镜所有终端仪器的某版本的 imap 文件 (见图 3 中的第 1 和 2 列)。综上所述, CRDS 主要基于四类文件展开: 三种层级式的映射/规则文件及最终映射推荐出的定标参考文件 (见图 3)。

表 3 CRDS 不同层级规则文件介绍

文件类型	文件说明	用途
pmap (pipeline context)	望远镜流水线层级规则文件	管理一个望远镜的所有终端仪器, 用于映射 imap 文件
imap (instrument context)	终端仪器层级规则文件	管理一个终端仪器的所有参考文件类型, 用于映射 rmap 文件
rmap (reference type mapping)	参考文件类型层级规则文件	管理某个参考文件类型的所有版本参考文件, 用于映射具体的参考文件

CRDS 的三类层级式映射/规则文件 (pmap/rmap/imap) 的结构及语法相同, 主要都包括两大部分: “头信息” (header) 部分和 “选择器” (selector) 部分 (见图 4)。“头信息”部分提供了描述该映射文件的相关信息, 包含文件来源、文件名称、望远镜名称、映射文件类型、验证文件完整性的信息等 (详见图 4 中 header 的内容)。其中, 非常重要的一个字段 (field) 是 “parkey”, 它表征的是 “选择器” 进行最佳参考文件查找时所使用的数据集参数 (通常为 FITS 文件中 header 的关键词, 或 JWST/Roman 数据模型名)。“选择器” 包括用于查找映射结果的匹配规则, 通常为嵌套结构, 可包括高层级的选择器和嵌套的子选择器。另外, 还可根据需要在 “头信息” 与 “选择器” 之间增加一些注释内容 (comment)。

关于规则文件的 “选择器”, pmap 和 imap 规则文件的 “选择器” 部分都较为简单 (见图 4 的第 1 列)。pmap 规则文件的 “parkey” 字段通常为 “INSTRUME”, 即表征终端仪器的关键词, 其 “选择器” 部分只需简单地匹配映射出不同的终端仪器对应的 imap 文件名。imap 规则文件的 “parkey” 字段通常为 “REFTYPE”, 即定标参考文件类型关键词, 其 “选择器” 部分也只需简单地匹配映射出该终端仪器的不同参考文件类型对应的 rmap 文件名。而 rmap 规则文件则较复杂 (见图 4 的第 2 列), “parkey” 字段的设置根据具体情况会包括多个关键词, 其 “选择器” 则通常为嵌套层级的选择器, 以匹配各类头文件的关键词。CRDS 规则文件的 “选择器” 可支持精确匹配、各种表达式匹配 (Match)、起始使用时间 (UseAfter) 选择、软件版本选择、时间最近选择、距离最近选择、范围选择等各类选择规则, 详见表 4。其中, Match 和 UseAfter 这两类选择器是使用最多的选择器 (见图 4 中的 selector 部分)。依赖于科学目标、观测策略、定标方案等多种因素, 不同望远镜的科学数据处理流水线对参考文件的选择可能会有不同的选择规则。另外, 科学家或其他用户若有较复杂的参考文件选择需求, 还可以通过自行修改本地端安装的 CRDS 路径中规则文件的内容来实现新选择规则的添加和使用。

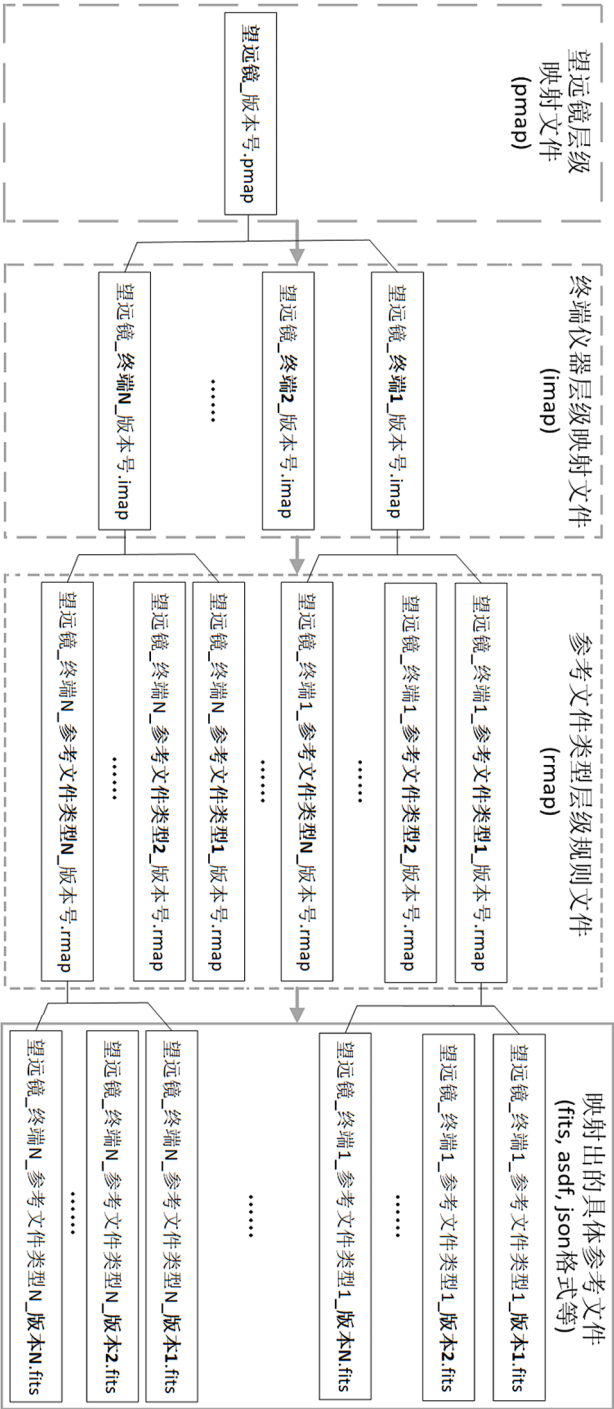


图 3 CRDS 规则文件的层级结构

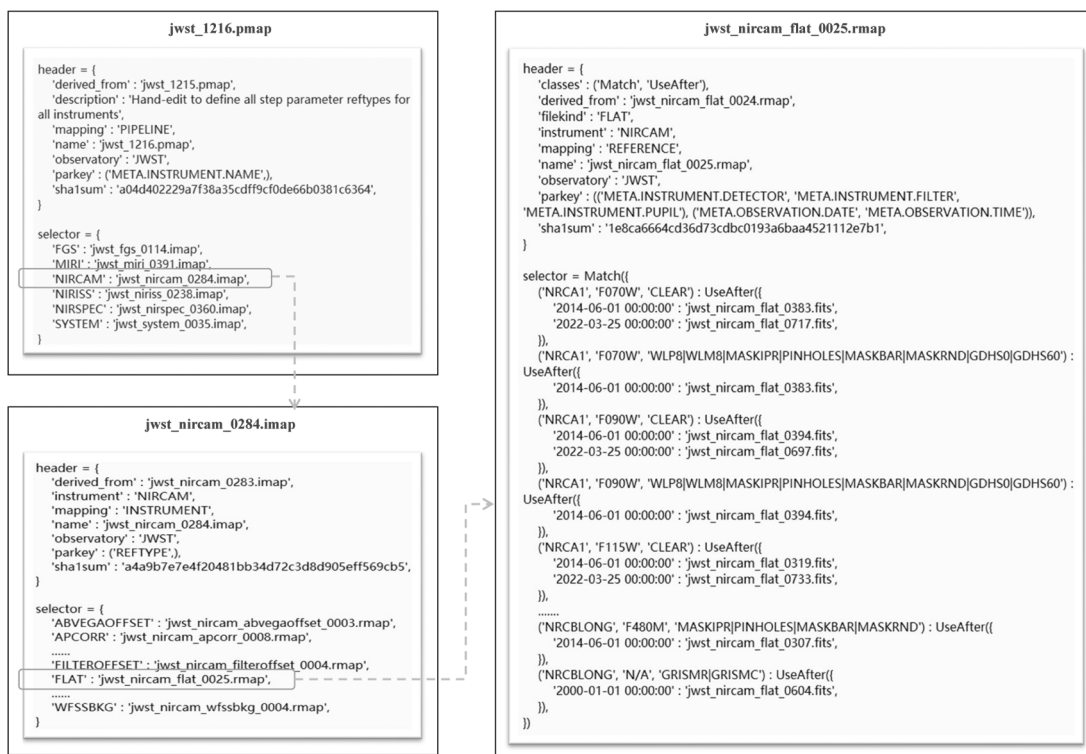


图 4 CRDS 的 pmap/imap/rmap 规则文件内容示例

表 4 CRDS 规则文件“选择器”的选择规则介绍

选择器	选择规则
匹配 (Match)	找出与科学数据头文件信息最匹配的元组, 支持精确匹配、枚举列表、通配符、正则表达式、文字表达式、关系表达式、范围表达、排除表达等
起始使用时间 (UseAfter)	找出早于 ($\leq$) 科学数据观测时间的最近时间的参考文件
软件版本选择 (SelectVersion)	使用流水线软件版本及各种关系表达式来进行选择, 即不同版本的科学数据处理软件可能会使用不同版本的参考文件
时间最近 (ClosestTime)	选择时间最近的
距离最近 (GeometricallyNearest)	选择距离最近的
范围选择 (Bracket)	用于返回两个选择结果

3.3 参考文件提交及检查

CRDS 能够推荐最佳参考文件的前提是存储和管理好提交的各类定标参考文件,并在提交参考文件时重构出相应的规则文件(见图 2 中的 CRDS 提交参考文件功能)。由于空间望远镜在轨定标会产生大量随时间更新的多版本多类型的参考文件,尤其是本底和暗电流这类探测器日常监测项,其更新频率为天量级,另外随着在轨后数据处理算法的更新,参考文件的生成也会不断更新,因此参考文件的提交会较为频繁,甚至存在大批量提交的情况。定标参考文件的提交,大体分为两类:增加一个之前没有的新类型的参考文件或替换已有的参考文件(如有更好的平场参考文件);增加已有类型的新版参考文件,用于处理某一个特定时间区间的科学观测数据。

如果单个参考文件出现问题就会影响大批科学数据的处理,所以参考文件的质量检查至关重要,即提交至 CRDS 的参考文件需要是有效的且绝对无误的。因此,提交定标参考文件到 CRDS 时,会进行一系列严格的校验/检查(见表 2 的 CRDS 验证检查功能),只有通过校验的参考文件才可以成功提交到 CRDS 服务器端,并触发对新版规则文件的重构。另外,重构出的新版规则文件也需经过检查验证后,才会存入 CRDS 服务器供科学数据处理使用。

参考文件及规则文件的校验检查主要是基于 CRDS 中定义好的后缀为 tpn 的限制文件(.tpn constraint definition files)开展。CRDS 特有的 tpn 限制文件定义了参考文件/规则文件(rmap 文件)所需的多方面的检查。对于 HST,其 tpn 文件几乎定义了其所有的 CRDS 检查,且是由原来的 CDBS 检查引入的,这也是 CRDS tpn 文件语法的来源。对于 JWST/Roman, tpn 检查用于进一步扩展 JWST/Roman 数据模型检查,例如,可增加“必需”检查选项、可检查矩阵维度、检查关键词之间相互关系、区分不同后端模块的可接受值等。tpn 文件的语法可支持显示(explicit)指令、合成(synthetic)指令、包括(include)指令、替换(replace)指令、约束(constraint)指令等^[43]。目前为止,使用最多的为约束指令,每行定义一个限制条件检查,通常包括 5 个字段:<名称>、<关键词类型>、<数据类型>、<存在条件>(即是否必须存在)、<值的约束条件>(逗号分隔开的一系列值,可使用 Python 表达式),5 个字段之间以空格分开。通过 CRDS 内部定义的 tpn 文件,可完成参考文件从头文件关键词到数据矩阵、表格等的非常全面的检查。

除了 CRDS 特有的 tpn 格式的检查,CRDS 检查校验(即 CRDS 的“certify”功能)还包括:JWST/Roman 数据模型检查、fitsverify 检查^①、表格行检查、FITS/ADSF/JSON/YAML 格式检查、rmap 规则文件更新检查、文件哈希值校验等。在提交参考文件或规则文件时,若存在问题而未通过检查,则会给出警告或报错;可根据警告或报错信息进行参考文件/规则文件的修改,以进一步完成提交。检查通过后的参考文件才会进行版本化重命名,并用于构建层级式规则文件,存储到 CRDS 服务器端以供科学用户使用。

^①<https://heasarc.gsfc.nasa.gov/docs/software/ftools/fitsverify/>

3.4 最佳参考文件推荐

CRDS 的核心功能是基于构建的规则文件, 为科学数据处理推荐所需的最佳定标参考文件, 即 CRDS 的 “bestrefs” 功能 (见图 2 及表 2 的 CRDS 推荐最佳参考文件功能)。单幅科学观测图像的定标通常需要非常多类型的定标参考文件, 这些参考文件因获取科学数据的终端仪器、观测模式等而异。因此, 作为定标数据产品的定标参考文件的头文件信息, 包含记录该参考文件相关信息的关键字, 包括实际观测的终端仪器、配置参数、参考文件类型、使用时间、制作时间等, 以及因数据类型不同而特有的关键字等, CRDS 会根据参考文件的关键字信息来重构生成新版规则文件, 再根据规则层级结构, 逐级映射分配出最佳定标参考文件。pmap 规则文件的更新则会在某个时间点, 为科学数据处理流水线设置相应的 CRDS 配置 (即当下有效的规则文件版本)。CRDS 通过读取待处理的科学数据的头文件信息, 并依据所选用的 pmap 文件, 首先根据 “INSTRUME” 关键字 (pmap 文件中 “parkey” 字段的内容) 的值, 匹配映射出 imap 文件 (包含在此 pmap 文件的 selector 中); 再依据所需的参考文件类型 “REFTYPE” (imap 文件中 “parkey” 字段的内容) 的值 (若未设置 “REFTYPE”, 则匹配该 imap 文件含括的所有类型的 rmap 文件), 映射匹配出 rmap 文件 (包含在此 imap 文件的 selector 中); 最后根据 rmap 文件中的选择规则及 “parkey” 字段, 推荐出最佳的定标参考文件 (见图 3 和图 4)。

对于 HST, 通过调用 CRDS 的 “bestrefs” 功能, 推荐出最佳的定标参考文件, 并将参考文件的文件名更新到科学数据的头文件中, 供后续科学数据处理使用。对于 JWST 和 Roman, CRDS 已完全集成到其科学数据处理流水线中^①, 流水线在运行时根据选用的 pmap 文件自动映射选择最佳的定标参考文件。所使用的 pmap 文件默认为 operational 状态的 pmap 文件, 也可根据需要选择旧版本的其他 pmap 文件。因此, 通常使用 JWST 数据发表文章时, 除了需注明所使用的 JWST 科学数据处理流水线的版本, 还需说明所用的 pmap 文件版本^[44, 45]和 CRDS 客户端软件版本, 因为这两者决定了数据处理所使用的定标参考文件。另外, 为便于追溯, 使用的 pmap 文件版本及最佳定标参考文件名, 最后会记录在处理后的科学数据的头文件关键词中。

使用最佳参考文件功能, 还需要有网络连接到 CRDS 服务器, 并配置好关键的环境变量, 包括 CRDS_SERVER_URL (CRDS 服务器端的 URL) 和 CRDS_PATH (CRDS 缓存的本地路径), 才可将所需的 CRDS 规则文件和参考文件下载到本地供数据处理时使用。这也使得用户可以方便查看以及灵活地修改规则文件, 满足数据处理中对定标参考文件使用的定制化需求。

3.5 科学数据再处理功能

定标参考文件和原始科学观测图像, 共同作为科学数据处理软件的重要输入, 决定了科学数据产品的质量。而科学数据不可避免地需要进行再处理, 其原因通常包括数据处理算法更新或定标参考文件的更新。对于空间望远镜, 因环境导致的仪器退化则需要周期性的在轨监测, 这会产生随时间变化的新版本定标参考文件; 同时在轨期间定标方案和定标产品生成

^①<https://github.com/spacetelescope/jwst>

算法的迭代更新,也会导致定标参考文件的更新提交。因此,定标参考文件需要频繁更新并提交至 CRDS,相应的规则文件也需要同步更新。CRDS 是基于规则文件来为科学数据推荐最佳定标参考文件的,一旦选择规则更新或增加了新的参考文件,势必会导致为相同的科学数据推荐出的参考文件发生变化,即意味着以往已处理过的科学数据可能需要重新处理。

再处理功能,也是 CRDS 的一个重要且新颖的功能。CRDS 的再处理功能通常与归档的数据库协同确定因新版参考文件的提交而需要重处理的科学数据集。当科学数据处理流水线选择了某新版 pmap 文件时,CRDS 会自动计算出由于规则文件的变化而需要重处理的科学数据,从而自动触发科学数据的再处理(见图 2 的 CRDS 再处理功能)。再处理功能会通过比较新/旧规则文件分配的参考文件,计算出受影响的科学数据集,CRDS 即可自动确定出已处理的科学数据集中哪些需要再处理,确认后,再处理系统会将日志和建议再处理的数据集存储下来,该结果可以通过 e-mail、客户端程序或网页界面获取。最后,由科学数据处理流水线运行科学家或科学数据的使用者来决定是否对受到影响的科学数据进行再处理,并执行数据再处理操作。

3.6 其他辅助功能

除了构建选择规则文件并根据层级化规则文件来分配最佳参考文件外,为便于规则文件和参考文件的管理,CRDS 还提供了多样化的辅助工具。这些辅助功能包括:CRDS 规则文件/参考文件/状态信息的下载和管理、约束检查和格式检查、文件之间的对比、内容查看、依赖查看等,详见表 2。这些功能和工具提升了 CRDS 系统的易用性与可管理性。

此外,为便于各项功能的实现,作为一套完备的系统,CRDS 有专门的网络服务器端。CRDS 的服务器端主要进行 CRDS 规则文件、参考文件及元数据的管理。不同的望远镜,各有其 CRDS 服务器端 URLs^{①②③}。CRDS 服务器端通过 JSONRPC 接口和 HTTP 接口提供了多种网络服务,包括参考文件/规则文件提交归档、最佳参考文件推荐、再处理功能、文件的传播下载等,来支持客户端软件的各类功能。图 5 展示了科学数据处理流水线从 CRDS 服务器端获取最佳定标参考文件的流程。

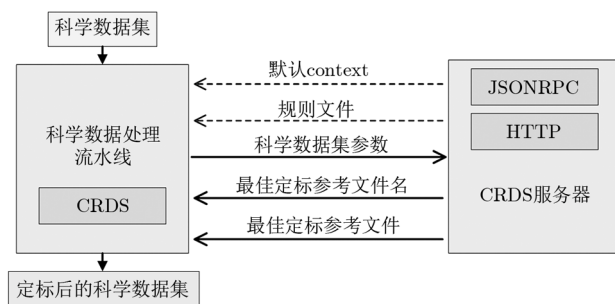


图 5 科学数据处理流水线从 CRDS 服务器端获取最佳参考文件示意图

^①<https://hst-crds.stsci.edu>

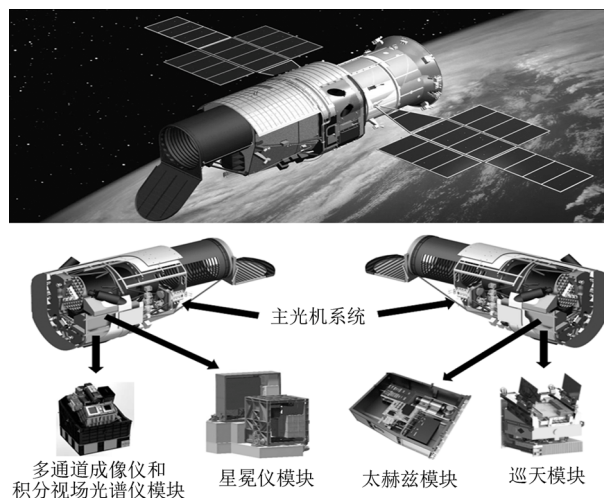
^②<https://jwst-crds.stsci.edu>

^③<https://roman-crds.stsci.edu>

4 CSST 定标参考数据推荐系统初步概述

4.1 CSST 介绍

CSST 是中国首个大型空间光学天文望远镜, 主镜口径 2 m, 有效视场 1.72 平方度, 未来将运行于近地轨道, 与空间站共轨飞行, 可与空间站对接维护升级, 计划在轨运行 10 年^[13]。巡天望远镜由五个后端模块(多色成像和无缝光谱巡天相机、多通道成像仪、积分视场光谱仪、系外行星成像星冕仪和太赫兹谱仪)组成(见图 6), 兼具大视场巡天观测、各类精细天文测量和深场/超深场观测等多种功能^[12, 46, 47]。CSST 有望在宇宙学、星系与活动星系核、银河系与近邻星系、恒星科学、系外行星与太阳系天体、天体测量和暂现源与变源等多个研究领域取得重大突破^①, 是中国空间天文的旗舰项目。



注: 图片来自中国科学院长春光学精密机械与物理研究所。

图 6 CSST 在轨运行示意图及其主光机系统和五个后端模块示意图

为实现“大视场、高像质、多波段”, CSST 配备的 5 台第一代观测终端仪器都极具特色。主巡天相机(main survey camera, MSC)的主焦面由 30 块 9000×9000 的 CCD(charge coupled device, 电子耦合器件)拼接而成, 总像素达到 25 亿, 其中 18 块探测器用于多波段成像, 12 块探测器用于无缝光谱观测, 天体定标组件采用与主焦面相同的 CCD^[48]。短波红外组件有 8 块 640×512 的探测器, 其中 4 块用于近红外波段成像, 4 块用于无缝光谱^[48]。多通道成像仪(multi channel imager, MCI)可开展紫外、可见光和近红外三个通道同时观测, 每个通道搭载可细分波段的滤光轮组件和与巡天模块一致的探测器^[48]。积分视场光谱仪(integral field spectrograph, IFS)采用基于改进型像切分器的视场积分单元, 由于空间观测不受大气视宁度限制, 该仪器能在整个光学波段实现高空间分辨成像光谱探测^[48]。系外行星成像星冕仪(cool planet imaging coronagraph, CPI-C)使用一种被称为

^①<http://nao.cas.cn/csst/>

星冕效应的技术,通过在恒星周围的视场中引入一个遮挡器或掩模来抑制恒星的亮度,从而使周围行星的微弱光线能够被观测到,此星冕仪主要用于在可见光至近红外波段对太阳系外行星进行观测^[48]。高灵敏度太赫兹探测模块(high sensitivity terahz detection module, HSTDM)在位于红外和微波之间的太赫兹频段进行探测,用于宽波段分子谱线的搜寻、银河系分子云的形成和演化、河外星系中性碳的精细观测、近邻宇宙中冷气体研究等^[48]。

4.2 CSST 定标数据推荐系统

CSST 科学数据处理系统是巡天空间望远镜数据处理的专门系统,主要任务有观测数据仿真、科学观测需求编排、数据处理。其中,数据处理是 CSST 科学数据处理系统的核心任务,需要对所有终端仪器的原始数据(0 级数据)进行处理,生成满足科学用户需求的 1 级和 2 级数据,最终为开展科学研究提供重要的观测基础数据^[49]。而 CSST 的诸多特点,包括后端模块多、观测模式多、对于巡天模块其多色成像和无缝光谱同时进行观测、大视场和拼接 CCD、科学目标多样等^[12],再结合复杂的空间环境等因素,都给其科学数据的定标带来了极大的困难,导致其定标数据极其复杂,也为定标参考数据的管理及分配带来了一定的挑战。因此,研发一套针对 CSST 的定标参考数据推荐系统(CSST calibration data system, CCDS),是科学数据处理系统研制期间的核心任务之一。该系统不仅需应对 CSST 在轨运行期间定标数据的复杂性,还需能够有序管理大量随时间变化的各类定标参考文件。通过制定合理的推荐规则,该系统将确保在空间望远镜复杂的在轨工作模式下,能够高效准确地将最佳定标参考文件推荐给科学数据处理流水线使用。

针对 CSST 的实际需求,定标产品可通过定标产品生成工具包和定标产品生成流水线两种方式获得。其中,定标产品生成工具包主要包括巡天模块中大多数监测频率不是很高的定标产品及精测模块的各类定标产品的生成,这些生成的定标产品在人工检查确保无误后才提交至 CCDS,即通过普通的网页或 Python API 的手动方式提交。而巡天模块探测器的日常监测项(如本底、暗电流、内部平场)因其较高的监测频率和较大的数据量,需开发为定标产品生成流水线,实现定标产品的自动生成、检查及自动提交 CCDS,且需要并行化进行,这就为 CCDS 提出了新的挑战。同时,定标产品生成流水线的部分产品的生成,需要调用和依赖其他定标产品,如暗电流产品依赖合并本底来生成,内部平场产品则依赖合并暗电流和合并本底来生成,这样相关的依赖逻辑又增加了 CCDS 系统的复杂性。图 7 显示了目前开发中的定标产品生成流水线与 CCDS 的交互示意图,可见,为满足大批量自动化定标产品生成及提交的需求,对 CCDS 的要求会更高。此外,考虑到 CSST 终端仪器及数据的多样性,其定标产品的数据类型也会非常多样,因此除了对 FITS/ADSF/JSON/YAML 这些参考数据格式的支持外,CCDS 还将扩展支持更多格式的参考文件,包括 TOML/PICKLE 格式的文件等。随着 CSST 科学数据处理系统的研制,及定标产品生成工具包/流水线的开发,对 CCDS 的需求会不断升级,使得 CCDS 更具拓展性。

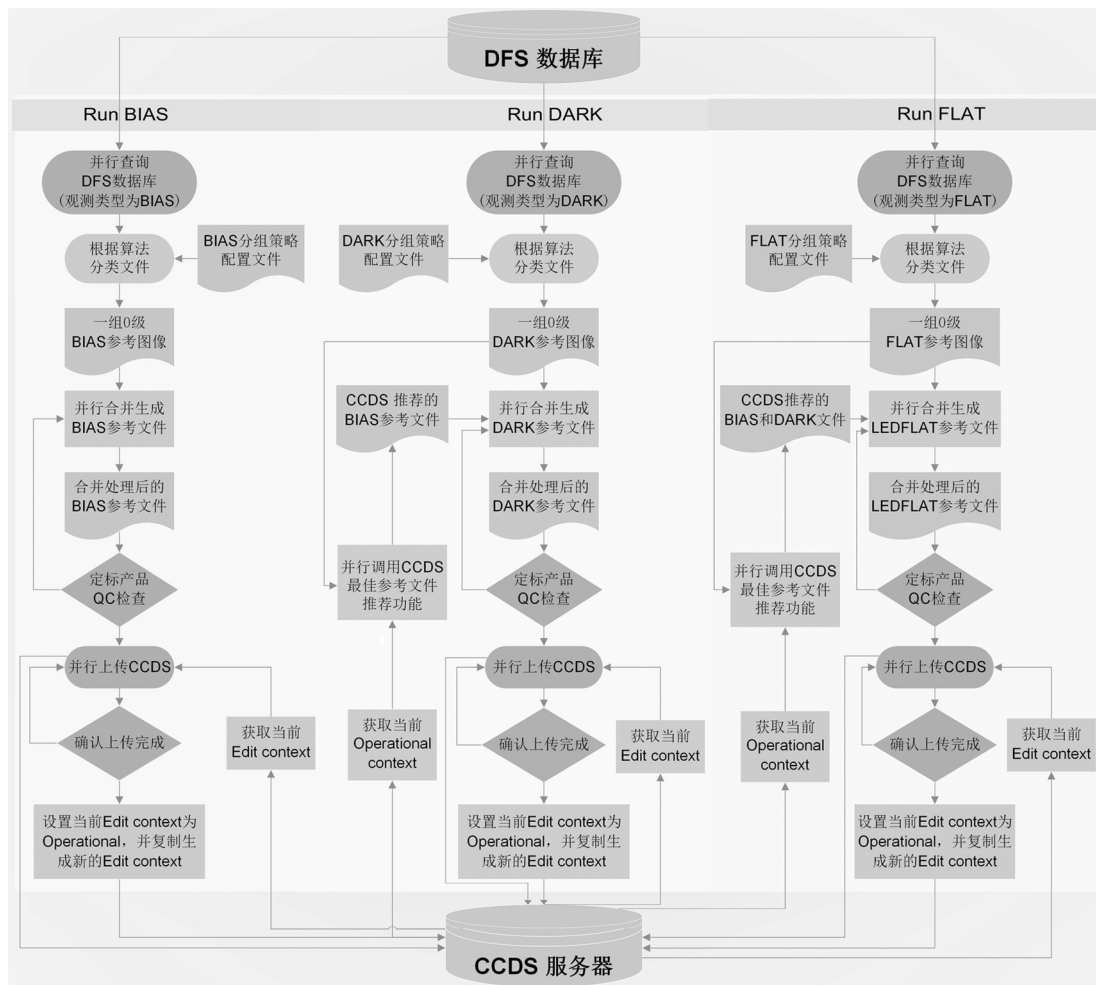


图7 CSST 巡天模块定标产品生成流水线与 CCDS 交互工作示意图

5 总结与展望

定标参考数据推荐系统在自动化天文数据处理中发挥着重要作用, 能够为不同时期不同终端仪器和观测模式的科学数据自动分配最佳的定标参考文件。经过国际上数十年的研究实践, 已逐步发展出了集定标参考数据的管理、规则制定和重构、自动化参考文件推荐等功能为一体的完善的综合性定标参考数据系统, 并在国际知名空间望远镜 JWST/HST 及计划中的罗曼空间望远镜上进行了实践, 可广泛灵活地应用于巡天及通用型望远镜的定标参考数据管理分配及自动化科学数据处理流水线的支撑上。

随着观测天文学的不断发展, 各类巡天及中大口径望远镜层出不穷, 对海量天文观测数据的自动化处理需求, 使得研究和搭建综合的定标参考数据管理和分配系统非常关键。特别

是类似于 CSST 的望远镜,它不仅配备了巡天模块,能够进行大天区面积的高分辨率多色成像和无缝光谱巡天,还集成了多种精测模块,用于开展各类精细天文观测。这种复杂性使得定标参考数据的管理和分配都极具挑战,急需一个完善的定标参考数据系统。随着计算机及多门交叉学科的不断发展,定标参考数据系统可借助更先进的网络技术和人工智能技术等不断优化完善,支撑大批量及超大容量参考文件的快速上传和管理,智能化选择规则的构建及参考文件推荐,使得整个天文数据处理的流程不断往高效和智能化方向发展,形成从观测数据到科学目标之间的智能化链路,最终达到最优化望远镜科学产出的目的。

致谢

本研究由中国载人航天工程巡天空间望远镜专项科学研究经费支持,课题编号为 CMS-CSST-2025-A19。感谢中国载人空间站工程巡天空间望远镜科学数据处理系统对本工作的支持,感谢审稿人的宝贵意见。

参考文献:

- [1] Harris D. *Astronomy: A Study of Celestial Objects*. New York: Murphy & Moore Publishing, 2022: 10
- [2] “中国学科及前沿领域发展战略研究(2021—2035)”项目组. 中国天文学 2035 发展战略. 北京: 科学出版社, 2023: 8
- [3] 黄佑然, 许傲傲, 唐玉华, 等. 实测天体物理学. 北京: 科学出版社, 1987: 3
- [4] 萧耐园, 宣焕灿. 图解天文学史. 南京: 南京大学出版社, 2012: 5
- [5] 程景全. 巨型望远镜时代: 现代光学天文望远镜. 南京: 南京大学出版社, 2023: 1
- [6] Djorgovski S G, Mahabal A A, Drake A J, et al. *Planets, Stars and Stellar Systems*. Dordrecht: Springer Reference, 2013: 223
- [7] Ferguson H C, Greenfield P, Axelrod T, et al. *astro2010: The Astronomy and Astrophysics Decadal Survey*, Washington: The National Academies Press, 2010: 15
- [8] Léna P, Rouan D, Lebrun F, et al. *Observational Astrophysics*. 3nd. Heidelberg: Springer-Verlag, 2012: 175
- [9] Freudling W, Romaniello M, Bramich D M, et al. *A&A*, 2013, 559: A96
- [10] Vernet J, Dekker H, D'Odorico S, et al. *A&A*, 2011, 536: A105
- [11] Zombeck M V. *Handbook of Space Astronomy and Astrophysics*. Third Edition. Cambridge: Cambridge University Press, 2007: 597
- [12] Zhan H. *ChSBu*, 2021, 66: 1290
- [13] Gu Y D. *ChSBu*, 2022, 37: 1031
- [14] Hands A D P, Ryden K A, Meredith N P, et al. *Space Weather*, 2018, 16: 1216
- [15] Doxsey R. *SpaceOps 2006 Conference*. Rome: AIAA, 2006: 5936
- [16] Yoshioka K, Miyoshi Y, Kurita S, et al. *Space Weather*, 2021, 19: e02611
- [17] Bohlin R C. *AJ*, 2016, 152: 60
- [18] <https://www.stsci.edu/hst/instrumentation/wfc3/calibration-plan>, 2025
- [19] <https://www.stsci.edu/hst/instrumentation/acs/calibration/calibration-plans>, 2025
- [20] Menzel M, Davis M, Parrish K, et al. *PASP*, 2023, 135: 058002
- [21] Rigby J, Perrin M, McElwain M, et al. *PASP*, 2023, 135: 048001
- [22] <https://www.stsci.edu/jwst/science-execution/approved-programs/calibration>, 2025
- [23] Freudling W, Zampieri S, Coccato L, et al. *A&A*, 2024, 681: A93
- [24] Gong Q, Bergkoetter M, Berrier J, et al. *Journal of Astronomical Telescopes, Instruments, and Systems*,

- 2020, 6: 045008
- [25] Bailey V P, Bendek E, Monacelli B, et al. Proc. SPIE, 2023, 12680: 126800T
 - [26] Scaramella R, Amiaux J, Mellier Y, et al. A&A, 2022, 662: A112
 - [27] Euclid Collaboration, Paterson K, Schirmer M, et al. A&A, 2023, 674: A172
 - [28] Greenfield P, Miller T. Astronomy and Computing, 2016, 16: 41
 - [29] Graessle D E, Evans I N, Glotfelty K, et al. SPIE, 2006, 6270: 62701X-1
 - [30] Lee W, Laher R, Fowler J W, et al. ASPC, 2005, 347: 594
 - [31] <https://www.stsci.edu/stsci/meetings/calhst/home.html>, 2025
 - [32] 李建, 崔辰州, 何勃亮, 等. 天文学进展, 2013, 31: 1
 - [33] Werner M W, Roelling T T, Low F J, et al. ApJS, 2004, 154: 1
 - [34] Fazio G G, Hora J L, Allen L E, et al. ApJS, 2004, 154: 10
 - [35] Hirst P, Cardenes R. Proc. SPIE, 2016, 9913: 99131E
 - [36] Jenkner H. European Southern Observatory Conference and Workshop Proceedings, 1988, 28: 357
 - [37] Cox C R, Tullos C. Proc. SPIE, 1993, 1945: 69
 - [38] Cox C R, Lubow S, Tullos C. Observatory Operations to Optimize Scientific Return, 1998, 3349: 218
 - [39] Swam M S, Lubow S, Hurt L. Astronomical Data Analysis Software and Systems, 2004, 314: 824
 - [40] Spergel D, Gehrels N, Baltay C, et al. <https://arxiv.org/abs/1503.03757>, 2025
 - [41] <https://readthedocs.org/projects/jwst-pipeline/downloads/pdf/latest/>, 2025
 - [42] https://roman.gsfc.nasa.gov/science/roes/Roman_WFI_Processing_v6.pdf, 2025
 - [43] https://hst-crds.stsci.edu/static/users_guide/index.html, CRDS User Manual, 2025
 - [44] Boyett K, Mascia S, Pentericci L, et al. ApJ, 2022, 940: L52
 - [45] Withers S, Muzzin A, Ravindranath S, et al. ApJ, 2023, 958: L14
 - [46] Gao M, Zhao G H, Gu Y D. Bull Chin Acad Sci, 2015, 30: 721
 - [47] Zhan H. Sci Sin-Phys Mech Astron, 2011, 41: 1441
 - [48] CSST科学团队. CSST科学白皮书, 2024: 16
 - [49] 载人空间站工程巡天空间望远镜科学数据处理系统. 软件系统设计说明, 2022: 5

Calibration Reference Data Recommendation Methods

ZHAO Bowei^{1,2}, WEI Shoulin³, REN Juanjuan¹, WANG Feng^{4,5},
LING Chenxiaoji¹, LIU Chao^{1,2}

(1. Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Faculty of Information Engineering And Automation, Kunming University of Science and Technology, Kunming 650500, China; 4. Center for Astrophysics and Great Bay Center of National Astronomical Data Center, Guangzhou University, Guangzhou 510006, China)

Abstract: The development and operation of automated astronomical data processing pipelines require the ability to automatically acquire the best calibration reference data across various categories, necessitating the emergence of a comprehensive Calibration Reference Data System. This paper reviews the various recommendation methods employed by different telescopes worldwide, and analyzes the strengths and weaknesses of each approach.

Additionally, it particularly introduces a new text-based rule recommendation strategy, along with a supporting comprehensive Calibration Reference Data System (CRDS). We also highlight the essential role and potential applications of this system in processing scientific data for the China Space Station survey Telescope (CSST), which is China's first large-scale space optical astronomical telescope. In conclusion, we provide insightful forecasts on the advancement prospects of this recommendation strategy.

Key words: astronomy data analysis; calibration reference data; space telescopes

《天文学进展》2025 年征订启事

《天文学进展》是天文学类中文核心期刊。刊物为季刊，2025 年本刊于 3、6、9、12 月 30 日出版，每期定价 100 元，全年 400 元（含邮寄费和包装费）。凡需订阅 2025 年《天文学进展》，请至所在地邮局进行订购。

统一刊号：CN 31-1340/P

邮发代号：4-819

欢迎订阅！谢谢支持！

《天文学进展》编辑部

2025 年 6 月

勘误声明

本刊 2025 年第 43 卷第 1 期的中文目录第一行，原文“恒星形成定律的研究现状”更正为“外太阳系小天体探测与研究进展”。对由此给读者、作者造成的不便深表歉意，敬请谅解。

《天文学进展》期刊编辑部

2025 年 6 月