

### 简讯

2024 年 12 月 5 日至 6 日，CSST 计算工作台模块团队在国家天文台进行了一次交流会议，主题聚焦于代码智能体的研究进展和需求讨论。CSST 科学数据处理系统的核心成员参与了此次会议。会议期间，CSST 计算工作台模块团队详细介绍了之江实验室天文计算研究中心在 CSST 智能体方面的规划与进展情况，主要包括代码转译、代码生成以及科学图表解读和问答能力在内的多项研究工作的最新动态。此外，之江实验室的人工智能工程师与国家天文台的天文学家就代码智能体在天文研究中的潜在应用进行了深入探讨。双方的交流富有成效，为未来的合作奠定了基础。本次会议不仅促进了双方的沟通与合作，还为代码智能体的研究方向提供了明确指导，为计算科学与天文学的交叉研究规划了具体的发展路径。

### 突出进展

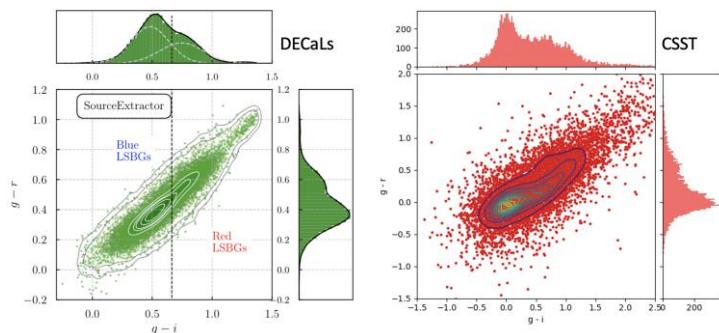
#### 数据处理系统在数据传输方面取得突出进展

近期，CSST 科学数据处理系统在数据输送方面取得了显著进展，成功实现了向 CSST 科学团队提供 0、1、2 级数据的目标。数据传输覆盖从数据需求申请到审核批准、数据准备、传输及记录的全流程。本月已完成两次数据输送任务，包括主巡天 50 平方度的仿真数据、多色成像 1 级数据产品（单次曝光星表和图像）以及 2 级合并星表，数据总量超过 5TB。这些数据主要用于测光红移和强透镜科学团队测试和研究。为满足不同需求，我们采用网络传输和硬盘拷贝等灵活方式，确保数据的安全、高效传输。这一全流程的成功试验意味着科学团队可以根据具体的科研需求，提前使用 CSST 的数据进行研究。不仅为 CSST 项目的早期科研成果产出打下了坚实基础，也标志着数据处理团队对科研支持方面取得重大进展。

#### 主巡天多色成像模块研制工作进展

多色成像流水线中的合并星表是指将单次曝光星表合并融合为多波段星表的功能。作为产生最终星表的核心模块，合并星表肩负着直接面对科学用户需求的重任。多色成像团队一直保持对合并星表模块的

不断迭代，取得了 data model 的定义、toml 文件生成和代码规范化等一系列进展，形成稳定的合并星表版本。在 C9.2 仿真数据的基础上完成处理并产生 366 个 brick 的合并星表数据。从中随机选取多个 brick 星表，与输入星表进行了位置匹配、颜色-颜色图比较等一系列对比，并且与 SDSS、DECaLS 等其它巡天进行了比较（见下图）。结果显示合并星表模块已经具备将单次曝光星表准确合并为多波段星表的能力。同时，多色成像团队还与之江实验室共同开展更加高效的合并星表的新版本开发，已初步编码实现。



#### CSST 计算工作台模块研制进展

CSST 科学数据处理平台的目的为了高效的管理天文科学数据，提高数据使用效率和数据处理分析速率等，这对整个 CSST 数据处理意义重大。截止目前，团队成员从平台需求分析、架构设计、环境搭建、数据处理和采集，以及平台测试等，已经具备了平台 V0.1 版本的发布准备。V0.1 版将在之江实验室发布，包括天文知识库，多波段融合，FAST 开发，CSST 开发，CSST 流水线管理等多用户多视角多功能，这为 CSST 数据处理奠定了坚实的基础，也为“CSST 多出成果、早出成果”拉开了帷幕。平台下一个版本节点在 12 月 31 日左右，将上数据管理，个人数据管理，数据分享等多个支持协作开发的数据管理功能。




**数据系统研制进展**

序号	模块名称	月度进展
1	多色成像	完成主巡天 v093 版本 50 平方度 1 级图像处理, 合并星表完成初步检查; 改进位置定标并加入容错; 改进流量定标消光对 nuv 和 u 波段影响, 计划于 50 平方度重处理时更新该功能; 初步完成合并图像 data model。
2	无缝光谱	进一步优化 COSMOS 天区的无缝光谱发射线诊断程序及光谱红移测量程序。初步完成 COSMOS 天区无缝光谱的天体分类程序, 并得到初步测试结果。持续参与流水线集成的相关工作。
3	数据流管理	完成主巡天流水线依赖的星表数据导入精度损失 Bug 的处理; 完成 MCI 的依赖的星表导入和查询; 完成 CCDS 中机器学习训练模型 onmx 文件类型的支持; 完成 Web 上支持流水线消息触发和联调。
4	流水线运行管理	开展 36000 条数据的 mbi 和 sls 测试; 讨论 DAG 设计, 初步完成 Scalebox 的 DAG 逻辑的流水线调度, 并进行测试; 完善 DAG 相关流水线运行管理界面。
5	多通道成像仪	完成仿真软件配置项测试说明文档闭环; 开展配置项安装和功能测试; 梳理软件需求规格说明书整理配置项测试内容; 开展 QC 流水线接入; 完成宇宙线扣除模块更新并上传, 通过单元测试。
6	积分视场光谱仪	完成仿真软件配置项测试说明的撰写和评审工作, 正在开展配置项测试工作; 完成参考文件工具包的开发工作; 根据测试数据更新部分算法。
7	星冕仪	参与星冕仪鉴定件指标测试工作, 完成图像拍摄和数据处理, 信噪比和对比度分析; 完成仿真实程序配置项测试; 对测试中发现的问题开展了分析和定位。
8	太赫兹	完成太赫兹模块曝光计算程序初步开发与调试, 程序已交付观测编排模块; 继续推进 2 级数据处理流水线开发和测试工作, OTF 处理模块计划接入 2 级数据处理流水线, 目前仍在调试; 梳理软件需求规格说明书, 为配置项测试做准备, 完成 C10 之后里程碑计划表。
9	天体测量	完成密集星场天区数据处理流水线优化; 继续开展点源定心算法优化; 梳理软件需求规格说明书, 为配置项测试做准备。
10	观测数据仿真软件	持续优化仿真软件, 解决了内存累积问题, 改正了星等计算不匹配, 添加了饱和亮星的 PSF 外插模块, 改正了导入亮星引起的部分模块计算越界。完成了 1000 平方度天区的类星体星表, 更新测试了星表导入接口。同时, 开展大星系贴图和仿真软件的配置项测试, 并填写测试报告。
11	数据可用性标记	利用 v093 版本数据对天区密度分布工具的模型验证测试, 并开展单元测试。
12	在轨定标	开展巡天模块内部和外部定标观测方案及内部定标观测所用时间的分析并迭代; 初步开发 CSST 观测天区选取辅助系统; 使用新版观测编排相关程序完成 12 个星场的 30 指向、5x5 指向等的指向坐标、观测时间、轨道信息等的编排, 为后续仿真等分析做准备; 迭代开展 7 个 offset 的不同曝光时间的仿真, 进一步分析不同波段的定标观测策略及精度; 开展定标参考数据定义版本更新; 提交 CCDS 用于无缝光谱目标探测的深度学习模型文件。
13	观测需求编排	完成主巡天仿真软件总体功能, 正在开展进一步优化; 开展 hstdm 编排程序接入网站; 整理前端 js 依赖, 统一不同页面的控件依赖版本, 更新 datetime 选取校验方式; 进一步优化定制界面及代码整理。
14	数据处理软件集成	通过 Redis 消息与 DFS-L1 和 Scalebox 实现多色成像 1 级流水线的自动化调度。
15	存储与计算	支持天文算子和 CSST 计算工作台在公有云的部署及发布; 利用低成本的数据湖查询方案代替 ClickHouse 等数据库提供的星表查询能力并通过验证, 正在开展与现有系统的集成工作; 开发自研的元数据管理系统以替代 datahub 作为存储系统的核心并部署至 z100 测试中。
16	天文算子	完成视觉基础模型预训练 300m 参数版本, 继续开展 600m 和 1b 参数版本训练工作。
17	CSST 计算工作台	完成科研工作台用户体系模块整体提测, 等待发布; 完成科研工作台 DFS_L1 功能融合 PRD 评审、技术评审, 进入开发阶段, 预计 12 月底完成。
18	数据融合	完成并发布“CSST 数据处理平台/多波段数据融合服务”网站年度版本, 集成天文数据查询检索、星表交叉匹配、任务智能分析计算等天文数据分析功能, 具备弹性分布式计算、数据高效索引与存储优化、复杂任务智能分析与执行优化等能力。

## 近期节点和计划

时间节点	计划任务安排	时间节点	计划任务安排
2024 年 9 月	完成 2 级数据定义会签 (已完成)	2024 年 12 月	完成《软件设计说明》评审
2024 年 10 月	完成《仿真软件配置项测试说明》评审 (已完成)	2025 年 12 月	基本完成 1000 平方度仿真
2024 年 10 月	完成数据系统 0、1 数据定义和定标数据定义版本更新	2025 年 1 月	完成编排软件 v1.0
2024 年 10 月	完成 50 平方度深场仿真	2025 年 1 月	完成《CSST 数据处理系统技术报告》
2024 年 11 月	完成《仿真软件配置项测试报告》评审	2025 年 1 月	CSST 科学数据处理系统 v0.8 版纳入受控库
2024 年 12 月	完成 C10 流水线集成测试		

编辑:

王海

3  
审核: 罗常春 2024.11.15

签发: 李维峰