

简 讯

AI 赋能 CSST 和 CSST 高级科学数据处理软件集成研讨会召开

2025 年 10 月 19 日-20 日，为期两天的“AI 赋能 CSST 和 CSST 高级科学数据处理软件集成”研讨会在上海天文台举行。本次会议由中国空间站巡天望远镜（CSST）科学数据系统和 CSST 科学工作联合中心共同组织，CSST 长三角地区科学中心承办，北京大学科学中心、国家天文台科学中心、长三角地区科学中心、粤港澳大湾区科学中心、数据科学中心及科学数据处理系统相关研究人员 30 多人参加。如何高效、精准地处理和分析 CSST 数据，挖掘最深刻的宇宙奥秘，具有重大挑战。与会专家围绕“AI 赋能 CSST 科学数据处理与科学研究的路径”核心议题，展开了多层次、跨学科的深度交流，深入讨论了 AI 与未来 CSST 数据处理分析深度融合的关键科学问题和 CSST 高级科学分析工具与科学数据处理系统的集成方式。经过充分研讨，决定未来分别围绕 CSST 多色测光图像，无缝光谱数据以及科学计算三个方向，分别组织团队进一步凝练 AI 相关的具体科学问题，协调推进 AI 赋能 CSST 相关的科学研究。同时通过讨论 CSST 高级数据处理软件集成，确定了初步的推进方案，为后期顺利集成奠定了基础。



最新进展

CSST 早期科学观测提案进展

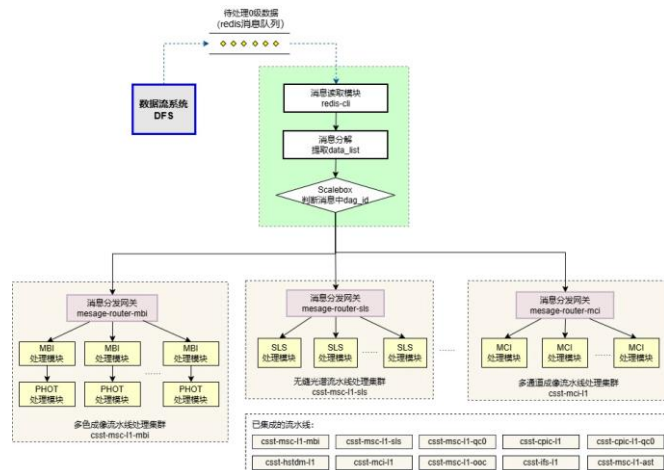
根据 2025 年 7 月底“CSST2025 年度科学年会”上关于早期科学的讨论和专家建议，为进一步凝练科学目标、统筹观测资源、提升早期观测效率和成果产出，联合中心于近期积极推动各研究领域和后端模块，针对第一轮提交的早期科学观测提案开展了阶段性优化工作。本次优化工作以“凝练科学目标、整合观测天区、优化仪器模式、提升观测效率”为总体原则，重点推进科学提案的合并与精简，旨在提升 CSST 早期科学的成果产出。相关学术活动主要包括：国家天文台举办的“恒星、银河系与近邻星系”研究方向优化会议；北京大学组织的“星系与活动星系核”领域

提案优化讨论会；云南大学牵头的“宇宙学”方向提案整合讨论；以及由 IFS、MCI、CPIC、HSTDMM 等后端模块组织的专题优化研讨。

通过系列会议和讨论，各团队围绕既定目标和原则对早期科学提案进行了系统梳理与合并，初步明确了 CSST 早期运行阶段的重点科学目标与观测方案，提出了发挥 CSST 核心优势的具体建议，并初步建立了科学工作小组的协同机制。这些工作为 CSST 在轨试观测阶段的科学规划打下了坚实基础，也为实现早期科学成果的最大化产出提供了有力支撑。

流水线运行管理模块研制进展

流水线运行管理模块团队围绕 CSST 科学数据处理系统流水线运行平台的核心目标，聚焦于突破性能瓶颈、保障系统稳定性，完成了一系列关键的技术优化与测试支持工作。首先，对核心技术进行架构优化与升级：（1）数据库层瓶颈攻关：针对底层数据库存在的唯一索引长度限制问题，深入分析了大数据量场景下的写入瓶颈，通过重新设计索引策略与表结构，成功解决了这一隐患，从底层保障了海量数据存储的可靠性与完整性。（2）消息流转机制重构：为高效处理日益增长的大数据量消息体，重新设计并实现了一套高吞吐、低延迟的消息流转方案。该方案优化了消息的分片、路由与处理流程，显著提升了平台的数据承载与传输效率。（3）平台处理逻辑迭代：基于上述底层与中间层的优化，同步对 Scalebox 平台的任务调度与处理逻辑进行了针对性升级。此次迭代确保了平台内核与新建消息流转机制的无缝协同，全面提升任务执行的稳定性和处理能力。其次，配合测试团队开展测试，对于测试过程中提出的各类问题，做到积极响应，快速定位根源，提供有效的技术解决方案，并及时进行热修复与版本更新，确保了集成测试的顺利进行。本轮技术优化和测试配合工作不仅支撑了项目关键节点的达成，还为项目后续大规模数据处理奠定了坚实基础，保障了项目顺利进行。



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 CTI 算法数据处理测试工作，在流水线中实现功能并对处理结果进行初步评估；二级流水线 mosaic 和强制测光模块完成单元测试和初步集成；针对配置项测试中的多个问题进行了算法修正。
2	无缝光谱	为增加 mosaic 抽谱而修改一级流水线相关的算法代码和数据定义，并通过集成测试；开展一级流水线配置项测试工作，准备测试数据样用例和完善测试说明文档；持续开展二级流水线各模块本地算法完善的工作。
3	数据流管理	完成 DFS 对齐新流水线调度和数据字段的修改工作；实现 csst-pipeline-cli。
4	流水线运行管理	针对数据库唯一索引对大数据量的限制，设计了全新的高吞吐消息流转方案，并同步升级了 Scalebox 平台内核；积极协同团队完成多条一级流水线的集成测试，并快速定位、解决测试过程中提出的各项问题，保障了测试进度。
5	多通道成像仪	新加入 CTI 改正功能完成并通过单元测试，正在开展集成测试；一级流水线数据处理集成测试全部完成，重启二级流水线单元测试工作，目前除图像叠加外均已通过测试；图像叠加由于采用 C 语言编写，正在协调新测试；开展一级流水线配置项测试，仿真数据生成、科学数据入库、编排文件入库均已测试通过；DAG 调用测试通过；正在编写配置项测试相关文档，测试用例所需数据已经部分生成，后续根据测试情况迭代。
6	积分视场光谱仪	完成一级流水线集成工作并通过集成测试，配置项测试人员成功开展测试；一级流水线对测试结果的正确性进行了分析；完成二级流水线单元测试，等待后续在之江平台开展集成。
7	星冕仪	完成新版编排程序的编写并和框架进行集成，可根据定义生成标准 json 文件；基本完成单元测试代码的编写；完成仿真程序的详细设计说明文档编写；配置项测试已经在公有云上跑通并完成说明文档，正在开展正式测试工作。
8	太赫兹	完成二级流水线单元测试，已基本具备公有云环境下集成测试的条件；开展编排程序更新工作和单元测试工作；在公有云环境下，继续推进一级流水线测试和配置项测试工作。
9	天体测量	完成一级流水线集成工作；完成二级流水线各个单元测试，模块之间的接口已经打通，进入一级流水线配置项测试阶段，编排文件及仿真图像入库均已成功，正在开展 DAG 测试工作。
10	观测数据仿真软件	完成 1000 平方度 W1 天区的 10 个指向的无缝光谱测试用仿真数据；完成 2*25 平方度 W5 天区 1 平方度强透镜仿真数据；开展软件配置项设计说明文档编写；开展暗电流模块升级，热像元模块更新工作。
11	数据可用性标记	完成天区密度识别模块单元测试并交付集成团队；正在开展地面测试大联测的工作梳理。
12	在轨定标	完成定标产品生成流水线/工具包相关文档编写工作：撰写软件功能说明文档、定标数据产品说明文档并已形成初稿；依据定标方案梳理了内部定标的具体观测策略，与编排团队进行了讨论，确定了跟编排的接口并分别生成了初步的内部定标观测策略和定标星场观测策略文件；补充 CCDS 新类型参考文件分别用于 MBI 和 SLS 的数据分类，其他多个类型参考文件根据 SLS 流水线及其他方面的测试的需求，也开展了更新工作。
13	观测需求编排	将各后端编排仿真软件单元融入 scheduler-standalone 框架，并基于此框架补充开展单元测试，解决代码库中 CI/CD Pipeline 出现的问题；开展配置项详细设计说明文档编写工作。
14	数据处理软件集成	初步完成多色成像 mosaic 和强制测光的数据产品定义和数据处理镜像；对齐 DFS1&2 的工作逻辑。
15	CSST 计算工作台	DFS2.0 增加二级流水线触发接口，完成开发工作并等待后续测试；完成星表查询接口增加元数据查询参数，关联查询元数据再查星表的开发工作；开展流水线重处理及日志查看功能。
16	数据融合	构建面向 JSP 图像融合分析过程的 JWST 图像元数据；研究利用 GPU 加速图像融合的 tracor_optimize 过程，相对 CPU 初步实现 50%以上的性能提升；实现无缝光谱参数测量模型中特征描述模块的 AstroDINO 的模型权重适配。

数据系统近期节点和计划

时间节点	计划任务安排	时间节点	计划任务安排
7 月 31 日	完成在轨定标方案评审出所评审（已完成）	11 月 30 日	完成 1 级流水线的全部开发和版本固化
8 月 31 日	基本完成 1000 平方度仿真（已完成）	11 月 30 日	完成数据处理软件新功能开发
9 月 30 日	完成多色成像 50 平方度深场仿真（已完成）	12 月 15 日	完成 2 级流水线的全部开发和版本固化
9 月 30 日	完成定标产品生成流水线和生成工具包固化（已完成）	12 月 31 日	完成数据处理软件 2 级流水线配置项测试
11 月 30 日	完成《软件设计说明》评审	12 月 31 日	完成编排软件配置项测试
11 月 30 日	完成数据处理软件 1 级流水线配置项测试	26 年 1 月	实现公有云环境下的部署，开展系统测试
11 月 30 日	完成无缝光谱 50 平方度宽场仿真		

编辑：王涛

审核：罗常青 刘飞

签发：刘锦峰