

CSST 科学数据处理系统月报

第 45 期

主办方：中国空间站巡天望远镜科学工作联合中心

时间：2024 年 11 月 11 日

新闻

《CSST 科学数据处理系统仿真软件配置项测试说明》通过评审

2024 年 10 月 31 日，中国科学院国家天文台在北京组织召开了载人空间站工程巡天空间望远镜科学数据处理系统软件仿真软件配置项测试说明评审会。参加会议的有来自中科院国家天文台、上海天文台、紫金山天文台、南京天光所和空间应用中心的专家，会议成立评审组。项目组成员首先做了《CSST 科学数据处理系统仿真软件配置项测试说明》（包括主巡天、多通道成像仪、积分视场光谱仪和星冕仪四个后端模块的仿真软件）的 4 个报告，主要内容包括仿真软件的软、硬件测试环境、测试数据、测试人员、测试规划、测试说明以及测试用例等。经过专家质询，与会专家认为观测数据仿真软件配置项测试说明中各模块仿真软件测试项和测试用例描述准确清晰，内容全面，测试方法正确有效，测试步骤具有可执行性，测试说明覆盖了需求规格说明的所有要求。同意通过评审。目前，项目组根据专家意见正在进行闭环。

简讯

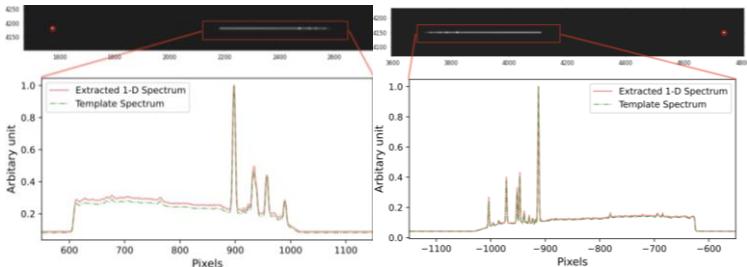
2024 年 10 月 14 日，由 CSST 联合中心举办的 CSST 科学数据处理系统-科学团队 2024 年研讨会在宁夏银川召开，会议旨在促进科学团队和数据系统的深入沟通融合。CSST 科学数据处理系统的多名核心研制人员参加了本次会议，汇报了关于数据处理系目前的整体情况，以及多色成像、无缝光谱、天体测量、积分视场光谱仪、仿真软件等模块的最新进展，并分享了一些核心算法的最新成果，最后双方进行了专题讨论。此次会议不仅加强了双方的交流合作，还加深了科学团队对 CSST 数据的认识，还为数据处理团队指明了相关改进方向，这对未来工作的顺利开展具有重大意义。

突出进展

在轨定标模块研制工作进展

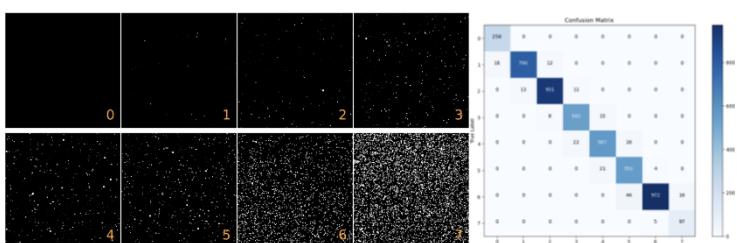
在轨波长定标是将无缝光谱数据的波长精度维持在所需范围内的重要步骤，是保障无缝光谱数据准确一致的基础。CSST 巡天模块无缝光谱包含 GU/GV/GI 三个波段，其中 GU 的波长定标是难点。GU 波段覆盖波长范围为 2500 埃至 4000 埃，而常用为无缝光谱进行波长定标的致密行星状星云或 Wolf-Rayet 星在该波长范围内尤其是靠近 2500 埃处发射线

数量较少，不便于批量快速的对该波段进行准确的波长定标。近期，在轨定标团队实现了行星状星云 PN G061.9+41.3 的 12 个无缝光谱 chips 的仿真，进行可同时对 GU/GV/GI 进行波长定标参考文件生成工具包的测试和优化开发。在对仿真数据进行定标测试分析的过程中进一步研究和改进了波长定标的算法：先将模板光谱与相应 chips 的响应卷积，得到初始参考谱，用来对仿真数据抽取的谱进行拟合匹配，直接得到色散信息，或之后再通过认证谱线得到色散信息。目前已经利用仿真数据测量了 12 块 chips 上 0 级像到 1 级谱的色散关系，生成了波长定标配置文件。下图为从仿真数据抽取的一维谱（红色）和模板谱（绿色），左右两图分别是左侧 2/5 和右侧 3/5 的谱（GU-01）



数据可用性标记模块研制工作进展

观测天区的密度分析与数据处理的方法密切相关。近期，数据可用性标记团队利用机器学习的方法，对主巡天多色成像和无缝光谱的图像分别进行密度自动识别判断：多色成像用残差神经网络对 CSST 仿真的多色成像图像数据进行密度分类，以便 pipeline 能针对不同密度的数据进行后续处理。从评测指标来看，除了最高密数据外（由于数据集过小，但也达到了 84%），不同类的分类准确率都超过了 90%。混淆矩阵显示错分样本都是被分到了相邻的两类，错分的原因主要源于数据集划分而非方法本身，这说明了我们方法的有效性；无缝光谱则利用 YOLO 目标检测算法和 CSST 仿真数据构建了一个无缝光谱目标检测模型。在验证集上，该模型零级像检测的查准率和查全率分别为 0.704 和 0.715，mAP50 为 0.698；谱线检测的查准率和查全率分别为 0.84 和 0.885，mAP50 为 0.891。下图为多色成像不同密度天区及结果示例。



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 nuv 定标评估并验证解决方案；针对流量定标问题进行不同消光天区数据检查；与仿真团队讨论并敲定深场区域；检查合并星表部分结果；开展 mosaic 相关讨论并制定初步开发计划；持续推进强制测光和模型测光研制工作；针对测光红移模块研制进行讨论并敲定时间节点，计划完成科学团队与数据系统对接的流程说明文档。
2	无缝光谱	完成了 COSMOS 天区无缝光谱和多色成像实测仿真数据的第一版生成，并对无缝光谱仿真数据进行了初步处理，同时完成了光谱探测率和信噪比等的初步统计分析。此外，还对无缝光谱发射线诊断程序及光谱红移测量程序做了进一步优化。目前，正积极开展对应的多色成像数据的处理工作，并探索无缝光谱与多色成像数据的协同处理方法。同时，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 v093 和 094 导入；完成外部星表数据 (trilegal_091,trilegal_093,trilegal_ooc_091) 导入和查询工作；完成 CCDS 参考文件删除优化；开展 CCDS 文件 S3 存储支持。
4	流水线运行管理	配合流水线集成进行 3 个版本星表合并流水线测试；开展流水线并行调度优化，支持自定义时间戳中增加标注。
5	多通道成像仪	完成 MCI 仿真软件配置项测试说明并开展配置项测试；配合流水线集成配置 QC 集成测试环境，开展并通过了初步测试，计划开展一级和 QC 流水线之间的集成测试调整；用新训练集生成 deepcr 训练模型用于宇宙线扣除；继续优化定标工具包，把各个功能拆分成单独的工具包。
6	积分视场光谱仪	完成 IFS 仿真软件配置项测试说明文档编写并通过评审；完成 0 级 QC 流水线接入；基于长光测试更新仿真软件的参考文件；基于 IFS 编排完成 NGC3359,20 次 300S 曝光的仿真。
7	星冕仪	完成星冕仪科学目标仿真和仪器仿真文章初稿并提交内审；完成仿真软件配置项测试说明并通过评审；完成仿真软件配置项测试用例准备工作，初步开展测试工作；正在开展数据处理流水线程序的整理和性能分析工作。
8	太赫兹	完成太赫兹模块科学数据处理一级流水线更新并通过单元测试与集成测试；开展太赫兹模块科学数据处理二级流水线更新，已有的模块通过了单元测试，最新版本的流水线通过集成测试；持续完善二级流水线。
9	天体测量	开展 5 平方度仿真数据的处理，一级流水线部分数据处理失败，正在根据处理结果优化软件；开展了 epsf 算法模型的构建工作。
10	观测数据仿真软件	持续优化仿真软件，更新了星表导入的信息提示接口，添加了探测器的串扰模块，完成新增模块的单元测试和 PEP8 代码格式规范，完成仿真软件配置项测试说明并通过评审，开展配置项测试工作；仿真星表方面完成了深场天区和 1000 平方度天区的恒星、星系星表，并调试星表仿真接口，开展相关仿真测试。
11	数据可用性标记	开发并完成利用机器学习的方法对主巡天多色成像和无缝光谱的图像分别进行自动识别的天区分析工具，计划开展该工具的整理工作并尽快完成单元测试和接入流水线。
12	在轨定标	完成 OOC 流水线，DFS 接口及单元测试，计划开展 QC 及集成测试；开展可观测性及定标星场选取约束研究，计算了十年的可观测天数；完成无缝光谱 GU 波段的波长定标参考文件，借助响应曲线使用了对连续谱和发射线整体拟合的方法，后续进行流量定标时将进一步检查其可靠性；无缝光谱 deepcr 宇宙线模型文件提交到 CCDS。
13	观测需求编排	新版主巡天编排仿真软件开发持续推进中；为方便新旧版程序切换，对网站部分程序进行了框架调整以便一边保留原有功能，一边将新功能同步接入；整理一般观测申请部署代码和镜像文件；对观测进展可视化进行前端及自适应显示优化、增加随时间动态切换观测天区功能。
14	数据处理软件集成	完成 v093/v094 仿真数据修改，完成入库并开始测试数据处理；讨论流水线 DAG 图的代码化和消息格式。
15	存储与计算	继续支持天文算子和 CSST 计算工作台的开发；计算引擎方面尝试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；存储引擎方面在 z100 集群上自建分布式存储已通过测试并投入使用，同时开始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。
16	天文算子	开展天文算子库主巡天星表合并流水线(csst-msc-l1-mbi-xcat)并完成上云改造。
17	CSST 计算工作台	完成科研工作台用户体系模块整体开发联调并进入测试阶段；新增科研工作台融合 DFS_L1 功能需求，正在进行需求调研，在线开发协作功能排期顺延；修改开发协作 PRD，对接 codespace。
18	数据融合	完成 AutoAstro 接口文档内部评审，初步构建交叉匹配和查询检索 tools 工具集；获取并生成超过 4 万张星表元数据表格并在此基础上实现用户星表元数据自动推导填充；相较传统人工构建流程节省 70% 左右元数据准备时间。

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		