

CSST 科学数据处理系统月报

第 31 期

主办方：中国空间站巡天望远镜科学工作联合中心

时间：2023 年 07 月 12 日



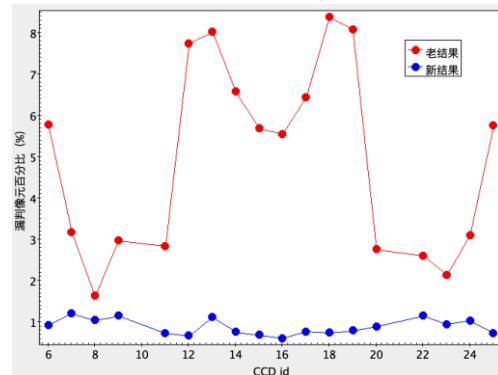
CSST 科学数据处理系统召开线下模块组长例会及中期评估进度会

2023 年 6 月 15 日-16 日，CSST 科学数据处理系统在上海天文台召开组长例会和中期评估进度会。会议对目前 Cycle7 阶段任务进行双周研制进度汇报，并对开展过程中遇到的问题进行了专项讨论，主要包括无缝光谱抽谱方案讨论、在轨定标相关讨论、望远镜指向的位置标定及新增观测模式的新需求讨论、50 平方度仿真及改版仿真软件功能模块需求讨论、一级流水线封装时间节点、流水线运行评估环境及二级流水线规划讨论、人机交互界面整合及网站服务开发讨论、数据处理系统软件知识产权问题讨论等；并对后期工作做出计划安排。同时开展了 CSST 科学数据处理系统研制项目中期评估进度会，对 4 月 14 日中期评估动员会布置的任务进行评估项的梳理、进度情况的汇报及相关问题的讨论，推动了中期评估准备工作的顺利开展。



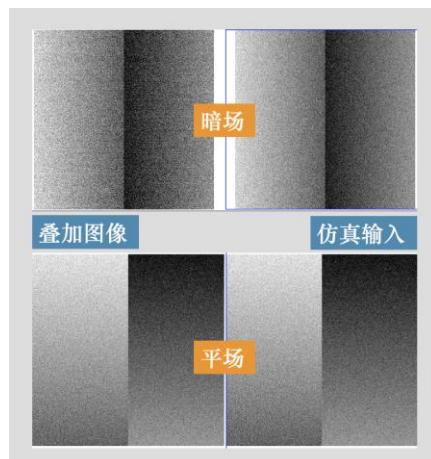
主巡天多色成像模块研制工作进展

团队研制人员根据 HST 太空望远镜的数据估计，判断 CSST 在巡天过程中将会受到严重的宇宙线影响，每块探测器在 150 秒的曝光过程中，有超过 20 万个像元会受到宇宙线的影响。为此开发了宇宙线探测模块。模块包括探测与填充功能，算法有 Lacosmic 和 DeepCR 两种选择。目前主巡天多色成像采用 DeepCR 深度学习算法，对宇宙线像元进行探测和标定。团队人员利用不同恒星密度的星场的仿真数据，对 23 号探测器数据开展了 DeepCR 深度学习训练，并将模型运用于仿真数据进行单图宇宙线探测。结果显示，对大于 3 个电子的宇宙线像元，DeepCR 漏判率在 3%-8% 不等。其中 8 号和 23 号探测器同属于 g 波段，漏判率最低；nuv 波段由于 PSF 轮廓与 g 波段差别较大，漏判率最高。针对此情况，团队人员分别对每个探测器都建立了各自的训练样本，并建立深度学习模型，得到 18 个对应不同成像探测器的宇宙线识别模型。新模型运用于今年 5 月发布的 2.0.0 版本仿真数据上，在不同探测器仿真图像上漏判率均能降低至 1.8% 及以下，提升效果明显。



在轨定标模块研制工作进展

在轨定标团队基于新仿真的带 overscan 区的 16 通道数据，开发了更新版的 CSST 巡天模块主焦面基本定标参考文件生成软件包（CSST_MS_CALIB），用于自动化生成部分基础且重要的定标参考数据，包括本底、暗电流、P-平场（即由原始内部平场图像叠加而来的平场参考图像）和增益（光子转移曲线法）。此外，该软件包使用 TOML 对输出数据格式进行规范；通过配置文件对坏像元和图像格式进行定义，并设置算法参数；其 overscan 模块可进行过扫区修正和图像格式转换；combine 模块可进行图像分区叠加和奇异值剔除；pipeline 模块则是小型流水线可用于快速获取叠加步骤前需要的图像；ptc 模块用于获得光子转移曲线和增益图像。由于其高度自动化的特点，该软件包可以作为定标产品生成流水线的核心算法软件包，根据（按照一定预设规则自动筛选的）给定的输入图像和基础信息，自动地生成基本定标参考文件。此外，也具有文件格式转换等附加功能，可以作为工具进行外部调用。相关代码已更新到 gitlab 上，并已完成单元测试。下图为使用该软件包进行的仿真测试结果：左为软件包生成的图像，右为仿真输入；上为暗电流图像，下为平场图像。




数据系统研制进展

序号	模块名称	月度进展
1	多色成像	完成对密集星场与 gaia 星表的匹配，初步判断匹配正确率高于 99%；对 Cycle6 仿真数据宇宙线探测成功率进行了对比；完成新流水线配置文件的更新，对每个探测器分别进行了训练学习，预期结果各波段接近 98%；仪器效应改正中快门、非线性和卫星轨迹本地测试通过。
2	无缝光谱	推进基于零级像抽谱工作；配合硬件测试团队开展相关工作；完成 2 平方度 Cycle6 数据中亮源的零级像认证、光谱提取工作；基于上述一级产品，已部署开展二级产品的验证工作；开发新的位置定标算法，无缝光谱区位置定标精度已有显著提升。
3	数据流管理	主巡天 2 级星表数据服务和检索优化；主巡天和精测模块的 1 级流水线运行支撑工作。
4	流水线运行管理	完善流水线运行管理界面，将主定时触发模块与各流水线模块的集成；完善流水线运行管理平台的时间戳管理；完成一级流水线的模块集成、运行测试、技术支持的工作。
5	多通道成像仪	根据一级流水线反馈迭代仿真数据并生成中期评估仿真数据；一级流水线完成 docker 封装并调试，配合镜像测试了 DFS 入库和 api 数据访问处理；准备 1 级数据定义 toml 文件并重新规划 QC 检查流程，需要后续代码修改并部署测试；二级流水线基于 LSST Plasticc 变源/暂现源模版库以及 MCI 相机参数和巡天策略，产生不同类型的变源/暂现源光变曲线样本，进而提取光变特征参数并进行分类。
6	积分视场光谱仪	在紫金山服务器上完整实现基于新的 0 级数据定义以及 docker 封装下和 dfs 交互的一级数据处理流水线的运行；完成 ifs 定位精度分析报告。
7	星冕仪	完善了使用多幅图像进行宇宙线识别的算法；完成新版本星冕仪数据处理流水线的单元测试和集成测试；结合观测编排和仿真程序生成了 24 小时观测仿真数据；完成星冕仪位置定标精度分析报告。
8	太赫兹	对 0 级数据仿真生成进行优化；进行 1 级流水线集成测试，进行 0 级数据反馈问题的修复并入库；完成 1 级数据定义文档更新。
9	天体测量	基于仿真图像完成单帧图像和天体测量 5 参数的对接，开展单帧图像相对天体测量精度分析和天体测量 5 参数精度分析工作；基于仿真图像近地天体的定位结果将近地天体轨道分类模块接入流水线。
10	观测数据仿真软件	更新了像场畸变模块、调整了仿真软件 WCS 的接口定义、完成了杂散光模块的集成和优化、升级了类星体星表。结合巡天策略，确定了 50 平方度的天区选择，并生成了对应天区的星系和恒星源表，用作 50 平方度图像观测仿真的输入星表。同时，团队开展了 CCD 实验室数据的测试分析工作，为后续程序改版提供参数接口。
11	数据可用性标记	1 级数据定义按照最新 Cycle6 数据产品基本完成 toml 文件及数据文件校对并通过单元测试；正在重新规划网页及大屏幕的布局；基本完成 QC0、QC1 流水线框架。
12	在轨定标	开发基于新仿真的带 overscan 的 16 通道数据的新版基本定标参考文件生成软件包并完成单元测试；BIAS/DARK/LEDFLAT/多色成像 SHUTTER 顺利重构规则文件并成功推荐最佳参考文件；无缝光谱配置文件测试 json 化中；开展定标星场仿真数据 PSF 重构分析；定标产品生成流水线完成 QC0 接入；完成 MCI、IFS、CPIC 定标参考文件的初步检查；MCI 研究并筛选了其中 300-400 颗具有 GAIA xp_sampled 光谱的恒星，正在剔除变源并计算 MCI 不同滤光片的标准星等；完成 IFS 基于 NGC6397 星团仿真数据的位置定标精度估计，正在筛选 NGC6397 中焦面 4 公共定标天区；正在进行 IFS 实验室测试光谱数据评价并与定标仿真预期比较。
13	观测需求编排	持续改进并整合编排仿真 python 包，以实现巡天观测编排；优化一般观测申请界面，支持后端更多个性化需求；修正警报解析 BUG 并考虑如何设置自动响应条件；调整观测进展可视化显示布局及人机交互，增加全局时间轴对显示内容进行控制。


近期节点和计划

2023年上半年Cycle7节点：

时间节点	计划任务安排	时间节点	计划任务安排
2023年2月底	与运控系统签订接口控制单	2023年7月底	模拟云环境下数据处理软件可运行
2023年3月底	与数据中心签订接口控制单	2023年7月底	数据处理系统软件 C7 原型
2023年3月底	配置项需求分析评审	2023年7月底	编排软件具备初步功能
2023年4月底	数据仿真软件 2.0.0 版本发布	2023年7月底	完成各模块一、二级流水线及定标流水线集成和运行
2023年5月底	配置项设计评审	2023年7月底	中期评估内审
2023年2-7月	C7 阶段各个软件配置项算法优化和功能完善	2023年7月底	收集鉴定件测试标定参数数据
2023年7月中	完成在轨定标方案评审		

编辑:

39

审核:

2023.7.10

签发:

刘继峰