

CSST 科学数据处理系统月报

第 12 期

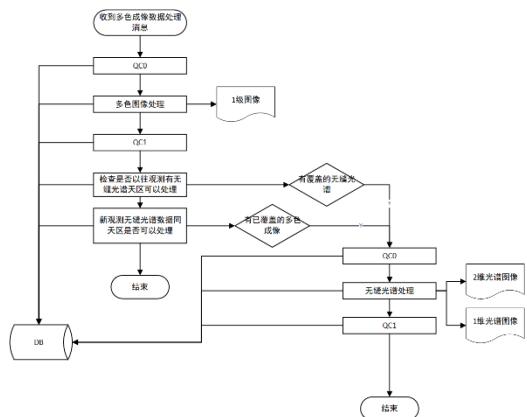
主办方：中国空间站巡天望远镜科学工作联合中心

时间：2021 年 10 月 10 日



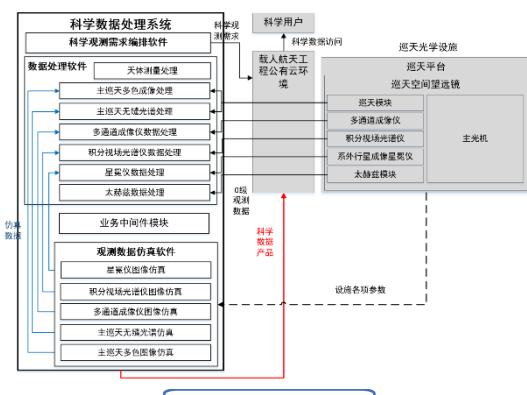
C4 流水线原型研发工作进展顺利

本月启动了C4流水线原型。在C3基础上加入0-1级无缝光谱数据处理部分，主要参加团队包括流水线运行管理、数据流管理、多色成像数据处理和无缝光谱数据处理模块。目前，已完成无缝光谱星表定义和无缝光谱流水线图。C4流水线原型构建的新流水线图见下图。



初步完成 CSST 科学数据处理系统-系统分析与设计方案

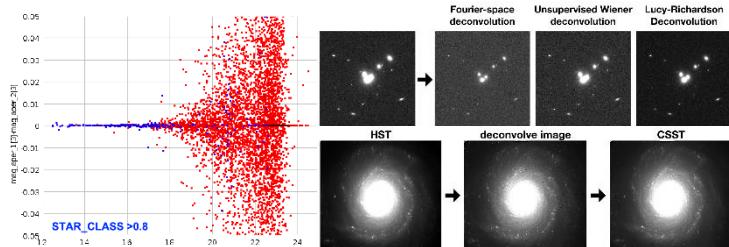
9月底,初步完成了CSST科学数据处理系统的系统分析与设计方案。本方案细化了实施方案中的系统设计,是项目研制过程中系统设计阶段重要的文档。它包括了整个科学数据处理系统中三大配置项(观测编排软件、仿真软件、数据处理软件)的各个功能描述和主要技术指标、系统的总体分析与设计以及系统内部各软件涵盖的流水线设计方案等。CSST科学数据处理系统整体流程图见下图。



多色成像模块研制工作进展

利用 BASS 数据, 检测了复杂天光背景下对流量定标精度的影响。BASS 巡天数据中有部分图像有较为复杂的天光背景, 利用算法可以获得天光背景并进行扣

除。多色成像团队通过对扣除天光背景前后的流量定标结果比较,发现背景对流量定标的结果影响不高于千分之一(1%) (见下图左)。结果表明,CSST 流水线中对天光背景的扣除功能可以在完成流量定标后,进行星表提取前再进行调用。



利用 HST 近邻大星系图像，构建 CSST 相近滤光片的仿真星系图像。CSST 巡天的一部分天区会受到近邻大星系的干扰，此类观测图像可能无法使用常规流水线进行处理。多色成像模块团队计划利用已有近邻大星系 HST 观测图像，使用退卷积算法模拟 CSST 观测图像。利用新开发的望远镜观测图像模拟程序模拟了视场为 180 x 180 角秒、曝光时间为 1000s 的 HST 观测图像，利用该模拟图像检验了 3 种退卷积算法，结果显示以 Lucy-Richardson 算法为优（见上图右上）；进一步利用 HST 实际观测的近邻大面积 face on 星系 NGC3147 图像进行退卷积，再利用 CSST 仿真团队提供的 PSF 构建了 CSST 观测系统下的近邻星系图像（见上图右下）。多色成像模块团队将继续扩大近邻大星系样本，并将样本库接入仿真图像结果，用于检验大星系数据处理的技术方案。

观测数据仿真模块研制工作进展

主巡天观测数据仿真团队于近日完成了仿真单元模块测试，并初步完成了《软件测试说明和测试报告—主巡天观测仿真软件》。本次单元模块测试历时近两个月，对软件的各个模块及外围软件进行了细致梳理，分解成光学系统、星表、无缝光谱、探测器、读出及其他共6部分，47个功能单元。按照测试计划要求，分离单元模块做成测试用例，针对正确性和代码覆盖率进行了系统地测试。

测试结果表明，各单元模块工作正常、输出结果符合预期。测试期间同时梳理了为数众多的全局和局域随机数种子的问题，这是导致前期模拟数据不准确的主要问题来源。主巡天观测仿真软件单元测试的顺利完成是该软件开发工作的里程碑，为下一步的集成测试和软件发布打下了坚实的基础。

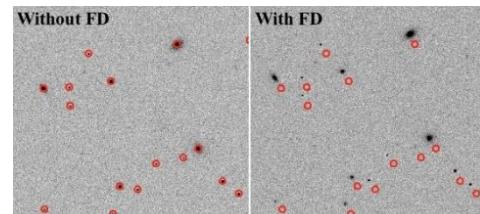


图: CCD#09 g 波段加入像场畸变前后的仿真图像比较。

数据系统研制进展

序号	模块名称	月度进展
1	多色成像	进行与无缝光谱相关的公共模块功能定义；汇总与仿真数据相关的问题；开展特殊图像（大星系，密集星场等）的处理分析和仿真需求。
2	无缝光谱	完成 Cycle4 阶段无缝光谱流水线原型架构设计；协助业务中间件和数据库模块初步完成无缝光谱接口；完成无缝光谱抽谱算法测试方案；面向仿真数据进行了目标源探测和 segmentation 生成测试。
3	数据流管理	开展在 10 节点的 Clickhouse 上模拟 10 倍的 Gaia DR3 的数据量的工作；对 Singularity 的使用学习并开展在 10 节点上使用 Singularity 部署 MySQL 数据库集群的工作。
4	流水线运行管理	继续进行流水线引擎在 TimeOut、异常处理等方面的原型验证，并考虑容错部分的设计。面向生产运行，启动流水线引擎的重构迭代，调整数据库设计，用 golang 重写执行引擎提高效率。
5	多通道成像仪	完成 MCI 数据系统从仿真到数据处理管线全链条测试，关键环节包括：优化仿真管线，生成 9K*9K 三通道仿真数据，包含各种仪器效应；基于仿真数据运行了单次曝光数据处理管线，初步实现全过程的数据处理；测试并优化超深场叠加算法。
6	积分视场光谱仪	和硬件团队沟通原始数据格式；分析多次曝光合成中计算资源和合成次数的关系；开展科学仿真模块的代码 review。
7	星冕仪	明确星冕仪观测的条件限制并分析各限制条件对目标可观测性的影响；进行初步的观测编排，完成 7 天的观测时间内星冕仪的编排样例；继续完善数据处理系统的设计方案；完成 1 级数据的头文件设计；梳理图像仿真和数据处理任务。
8	太赫兹	进一步梳理模块数据处理流程，完成了结合卫星运行轨道的点源目标频谱数据模拟生成，打通了点源目标同太赫兹模块运行轨道之间的仿真数据联通壁垒；完成了模块观测任务在多约束条件下观测任务编排的理论建模，在此基础上开展观测任务编排仿真工作并获得成效。
9	天体测量	优化了图像处理中连通域算法中背景扣减的策略；完成 CSST 科学数据处理系统天体测量数据处理的系统设计分析方案；与外协单位开展外协合同指标和周期论证工作，形成了初步合同文稿。
10	观测数据仿真软件	开展独立单元模块的覆盖度测试，整理完善单元测试报告；系统测试检查了光学系统模拟、半解析星表检测、无缝光谱仿真、探测器效应、CCD 读出等关键部分，为仿真软件的后续升级和维护提供有效保证；与长光所空间光学研究团队研讨空间站望远镜的杂光建模和计算，分析讨论了目前杂光模型的精度和误差来源，为下阶段仿真软件添加杂散光模块做好准备。
11	数据可用性标记	初步完成多色成像 QC1 的可用性标记评估参数列表初稿，正在搭建 QC1 数据可用性标记的框架并按照公共模块的定义重新架构 QCO 的框架。
12	在轨定标	与仿真团队进行主巡天定标星场仿真需求方面的沟通，初步完成了一个指向的定标星场的仿真（150 秒曝光和 30 秒曝光），并与多色成像和无缝光谱数据处理团队沟通了后续的数据处理安排，进行仿真数据的检查及处理；初步规划和搭建在轨定标策略方案书的框架，待确认后计划于下个月启动方案书的撰写；参与器件级测试及系统级测试的地面测试大纲制定工作，继续完善测试项目、测试指标需求。
13	观测需求编排	对系统框架进行整体调整，实现所有代码在 csst.planning package 中的系统调试与网站部署；依托 unittest+nosetests 实现代码测试与测试覆盖度评估框架，各模块未来只需要写 unittest 测试代码即可完成单元测试及评估；持续推进各子模块开发：一般观测申请部分修改了观测源数据模型及相关界面；警报系统初步部署 KAFKA 的消息系统测试环境并进行消息读写实验；编排需求生成部分导入 IFS、MCI 观测编排输入，准备开始进行在线编排生成。

近期节点和计划

2021 年下半年 Cycle4 节点：

时间节点	计划任务安排	时间节点	计划任务安排
8 月中	完成多色成像流水线原型评估	8-12 月	各数据处理模块流水线搭建，实现无缝光谱流水线原型
8 月底	完成 0 级数据结构定义文档评审	8-12 月	公共模块的定义和开发
9 月	完成阿里云环境测试	8-12 月	开展系统软件设计
9 月底	完成系统分析设计方案评审	10 月	完成系统功能需求评审
8-10 月	完成 1-2 级数据定义并评审	12 月	完成在轨定标策略方案书评审
8-10 月	完成在轨定标流水线搭建	12 月	完成仿真软件 V1.0 版本的发布
8-12 月	完成科学需求编排软件原型搭建		