

CSST 科学数据处理系统月报

第 21 期

主办方：中国空间站巡天望远镜科学工作联合中心

时间：2022 年 8 月 15 日

简讯

CSST 科学数据处理系统 Cycle5 工作总结会在云南丽江顺利召开

7 月 22 日-24 日，CSST 科学数据处理系统在云南丽江召开了 Cycle5 阶段的工作总结会议。本次会议由中国科学院国家天文台主办，来自中国科学院国家天文台、上海天文台、紫金山天文台、南京天文光学技术研究所、计算机网络信息中心、中山大学、昆明理工大学、北京师范大学等多家单位的科学数据处理系统软件模块组长和技术骨干共约 35 人参加了本次会议。应用总体相关人员也通过线上线下方式参会。此次会议中，科学数据处理系统各模块组长对 Cycle5 阶段的工作进行总结并介绍了 Cycle6 阶段工作计划。会议还针对若干重要技术议题展开了专项讨论并提出解决方案。会后，对各个专项讨论进行了要点整理，并对 Cycle6 的工作计划进行了总体梳理，为项目的顺利开展奠定了良好基础。



CSST 科学数据仿真交流会在山东青岛召开

8 月 6 日-9 日，CSST 科学数据仿真交流会在山东青岛顺利召开。本次会议由中国科学院紫金山天文台主办，邀请了来自国家天文台、上海天文台、长光所、上海交大、云南大学以及浙江大学等一线科研团队的科学家，就仿真数据的生成和处理需求开展广泛的交流讨论，会议形式以特邀报告和讨论为主，共设置 4 个议题。目前，观测数据仿真软件已完成第一版发布，仿真软件团队和科学处理软件一级数据流水线分析团队正在迭代对接，分析小规模模拟天区的图像。仿真数据的上下游用户就数据本身进行交流研讨，同时规划部署下一阶段任务。

突出进展

仿真数据软件模块研制工作进展

主巡天仿真软件 1.0.0 版自 4 月发布以来，团队成员收集整理用户关于软件使用和数据问题的反馈，针

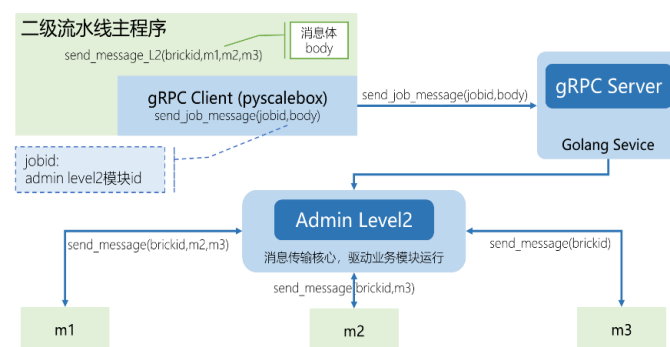
对相关问题更新升级了对应模块，并于近日完成了新版软件（1.1.0 版）在 gitlab 上的部署发布。在这一版软件发布中，规范了软件开发和问题反馈的管理平台，改正了天测模块调用错误和输出星表选项问题，调整了成像和光谱仿真两部分泊松噪音的添加模型，优化了仿真执行的记录输出和部分库函数的版本依赖，升级了天体星等转化模块，提高了仿真计算效率，添加了 16 通道增益的模块开关和截断星等计算器模块。基于 1.1.0 版仿真软件，同步完成了仿真数据的更新和线上版本的维护升级，并在近期通过线下研讨的方式，与数据处理团队就新版数据的测试分析工作开展充分讨论。

同时，仿真团队积极组织开发星系贴图模块，开展多组不同工况光学 PSF 数据的仿真工作和不同工况 PSF 数据植入模块，对比分析不同星系光谱模型的模拟数据，对不同杂散光模型开展对比测试，分析杂散光对无缝光谱仿真的影响，为下一版主程序的迭代升级做准备。

流水线运行管理模块研制工作进展

流水线运行管理团队于本月完成了流水线主程序与流水线运行管理模块 python 接口的设计与实现，支持跨系统的基于 gRPC 消息传输；设计统一的流水线运行管理框架支持一级、二级流水线的运行调度，其中一级流水线按需支持单模块运行和控制模块+运行模块两种模式，二级流水线基于消息转发主模块驱动业务模块操作。二级流水线运行框架初步完成三个部分：（1）嵌入二级流水线主程序的 gRPC 客户端（pyscalebox），与流水线运行管理模块通讯；（2）流水线运行管理 gRPC Server 端，实现二级流水线消息队列接口；（3）主模块 Admin Level2，是二级流水线运行框架中的消息传输核心，驱动所有业务模块有序运行，每个业务模块接收到主模块消息，处理完成后发消息返回给主模块。

通过对比 gRPC 的 protobuf 的二进制数据流和传统的 json 数据流，gRPC 的 protobuf 无论从体积还是编码格式方面均有明显的优势。采用高性能 gRPC 进行消息传输，为流水线快速调度及时响应奠定了基础。



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 Cycle5 的 2 平方度数据（添加天测信息）单次曝光 pipeline 全流程处理；针对数据的畸变解算效果不好的情况，重新开发了针对 CSST 数据的畸变拟合；完成了仪器效应、定标等步骤的结果简单检验，协助仿真发现 throughput 问题。
2	无缝光谱	对不同类型天体（恒星、星系、类星体）的抽谱结果进行了初步评估；制定了 Cycle6 研制任务具体实施方案，补充了对地面测试的需求内容。
3	数据流管理	测试数据库 Clickhouse 和 Yugabyte 星表导入性能；使用紫金山实验室提供的 3 节点，搭建 DFS 基础测试环境；在昆明理工大学和广州大学分别完成基于云的 gitlab autorunner，开展自动测试；基于 100GB Infiniband，开始测试 BeeGFS 性能，为大数据访问进一步加强技术储备；在星表合并取得一定进展。
4	流水线运行管理	完成流水线主程序与流水线运行系统的 python 接口的设计与实现，支持跨系统的消息发送；设计统一框架支持一级、二级流水线的运行调度，其中二级流水线基于消息转发主模块驱动模块运行，一级流水线按需支持单模块运行、控制模块+运行模块两种模式。
5	多通道成像仪	完成 Cycle6 计划书；开展三通道宇宙线扣除测试。
6	积分视场光谱仪	完成 IFS 仿真和数据处理软件 Cycle5 任务总结和 Cycle6 任务细化；开展 IFS 数据处理软件核心算法的文档撰写及内部评估过程。
7	星冕仪	参加 Cycle5 阶段工作总结会并对模块进展进行了汇报；对地面测试需求和 1 级数据文档进行了更新。
8	太赫兹	开展太赫兹模块数据处理依赖的第三方库从 windows 环境移植到 linux 环境方面的研究，解决 SOFA 标准库移植到 linux 供 python 调用的问题。
9	天体测量	单帧图像天体测量，搭建了包括参考星表构造及调用、参考星坐标转换、图像预处理、目标检测、天文定位等多个模块的数据处理 pipeline。
10	观测数据仿真软件	完成新版 1.1.0 软件在 gitlab 上的部署发布；同步完成仿真数据的更新和线上版本的维护升级；组织开发星系贴图模块，开展多组不同工况光学 PSF 数据的仿真工作和不同工况 PSF 数据植入模块；对不同杂散光模型开展对比测试，分析杂散光对无缝光谱仿真的影响，为下一版软件迭代升级做准备。
11	数据可用性标记	精密稳像小天区分析正在进行全巡天的分析，预计将于本月下旬得到第一次的分析结果；展示平台界面正在修改完善；基本模块代码正在进行单元测试的编写工作。
12	在轨定标	使用密集星场测试数据，基于星表的积累合并，开发测光平场的精细化测量和拟合的代码；开展早期星定标星场的选取；开展 1 级数据定义中的定标参考数据的定义和讨论；开展 CRDS 自动生成新版规则文件功能的研究和测试；完成新一版的定标星场仿真数据，正在生成定标参考文件，并添加 CRDS 所需的关键词，以进行 CRDS 测试。
13	观测需求编排	进一步梳理编排流程及制定 Cycle6 计划；设计插入替换编排的操作流程并开始编写相关功能；调研在线可视化方案，并基于 TLE 数据利用 WWT Web Engine 模拟空间站在轨位置。

近期节点和计划

2022 年下半年 Cycle6 节点：

时间节点	计划任务安排	时间节点	计划任务安排
22 年 7 月中旬	系统分析设计说明外审闭环	22 年 9 月	协调与数据中心的外部接口
22 年 7 月中旬	Cycle5 任务总结/Cycle6 计划	22 年 10 月	配置项需求规格说明书评审
22 年 7 月中旬	V1.1.0 仿真软件发布	22 年 10 月	协调与运控系统的外部接口
22 年 7 月-12 月	不同探测器地面测试	22 年 12 月	配置项软件设计评审
22 年 7 月-12 月	观测需求编排外部接口定义	22 年 12 月	实现所有模块统一程序框架并完成主要功能封装
22 年 7 月-12 月	开展算法科学验证	22 年 12 月	二级数据流水线应用接口发布
22 年 8 月底	一级数据定义	22 年 12 月	Cycle6 任务总结