

CSST 科学数据处理系统月报

第 39 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 4 月 15 日



CSST 科学数据处理系统召开线下组长会

2024 年 3 月 31 日-4 月 3 日，CSST 科学数据处理系统线下组长会在广州召开。本次会议首先各模块组长对 Cycle9 阶段近两周的研制进展进行汇报，之后会议针对项目研制过程中遇到的问题进行了相关专项讨论，主要包括观测数据仿真软件的软件发布、软件测试组织及 1000 平方度数据处理的讨论；图像重采样和合并的工具相关讨论；编排软件/仿真软件/数据处理软件/在轨定标工具包的单元测试讨论；数据处理软件界面、编排软件界面、QC 界面的问题讨论；0 级数据定义、定标参考数据定义的讨论；云计算流水线设计方案及相关讨论等。本次会议对仿真、测试等本阶段重要工作节点任务进行了充分交流讨论，有助于更好地开展后续工作。本次会上还完成了《CSST 数据处理系统 0、1 级数据数据输入需求和数据结构设计说明》两个文档（后面简称 0、1 级数据定义）的内部会签工作。



0、1 级数据定义文档完成内部会签

CSST 数据处理系统完成第三版 0 级数据定义和第一版 1 级数据定义并在 3 月组长会完成内部会签，纳入 CSST 科学数据处理系统配置管理功能基线受控库。

0 级数据是指未经过仪器效应扣除和数据处理的原始观测数据，作为科学数据处理流水线的输入数据。0 级数据的完整性和准确性对后续的科学数据处理会产生直接影响。0 级数据采用 FITS 格式，其中的全局头文件记录了观测目标以及曝光开始和结束时刻的望远镜信息，图像扩展中的头文件记录了后端模块信息，数据部分存储了原始观测图像。各模块 0 级数据的文件设计，命名，存储以及具体的数据结构都在 0 级数据定义文件里进行了规范。文件中同时定义了观测的唯一识别号 obsid (11 位)=模块编号 (1 位)+观测类型 (2 位)+曝光编号 (8 位)。

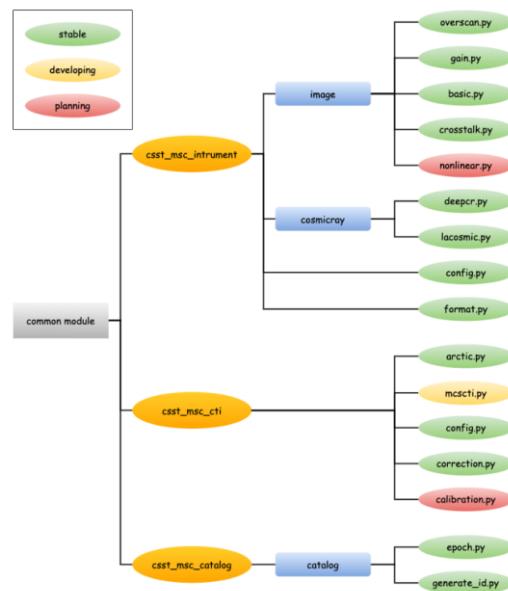
1 级数据是指通过数据处理流水线产生的 1 级科学数据和 1 级数据产品，对 CSST 的 0 级数据产品通过 1 级流水线数据提取和处理，包括对探测器的图像进行仪器效应改正，背景光去除和宇宙线扣除等处理，并进行一系列的定标处理，从而形成 1 级科学数据产品。1 级数据定义文档包括了各模块的 1 级数据文件设计、命名和存储规范；接口设计示意图以及数据元素示例。

此次系统内部会签的 0、1 级数据定义会作为 Cycle9 模拟数据和数据产品的接口标准，为流水线顺利开发迭代奠定了坚实的基础。

数据处理软件主巡天公共模块研制工作进展

主巡天公共模块目前包括 3 大功能模块：仪器效应改正、CTI 和参考星表改正。代码仓库分别为 csst_msc_instrument、csst_msc_cti 和 csst_msc_catalog。其中仪器效应改正主要指探测器效应的图像改正如过扫区、增益、本底、暗场、串扰、宇宙线等效应的改正；CTI 是 CCD 特有的较为复杂的探测器效应；参考星表改正主要指历元改正、ID 生成等等功能。目前已经基本完成其主要结构的构建（见下图）。根据设计要求，公共模块的功能子模块分别对应流水线的子函数，按流水线要求进行设计，方便流水线调用。csst_msc_instrument 和 csst_msc_catalog 已经完成基本功能代码并完成单元测试。可以保证在 Cycle9 阶段主巡天数据处理流水线调用这两大公共模块，完成流水线集成。

公共模块的使用不仅减少了不同流水线在处理同样功能的时候出现差异；而且也减少了开发的工作量，我们未来还会有更多的共用处理功能加入，保障流水线顺利高效地运行。



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 0 级和 1 级数据 fits 模板和 toml 文件的闭环，0 级数据和仿真 fits 文件正在对齐；完成仪器效应改正的初步更新：使用仿真最新 fits 文件，完成公共模块调用；完成天光背景扣除功能完成，已经按照 1 级数据定义输出完成对齐；仿真星表改版完成，并加入了多波段星等和位置误差；photutils 孔径和 kron 测光已完成初步封装；强制测光流水线基本完成，部分功能等待优化和开展单元测试。
2	无缝光谱	进一步优化无缝光谱天体分类程序并改善星系与类星体红移的拟合算法；对 3.0 内测 α 版无缝光谱的 PSF 进行初步测试；开始将基于零级像的抽谱程序集成到流水线中；协同完成 0/1 级数据定义工作。
3	数据流管理	初步完成 DFS 后端模块重构同时支持使用 WEB 和 GRPC 的接口访问；数据视图的 WEB 框架构建，实现了 0 级数据的展示和重处理的消息发送。
4	流水线运行管理	主巡天的流水线运行与数据流管理集成方案确定，完成相关链接页面；流水线运行平台数据库功能完善；完成三级监控信息采集功能。
5	多通道成像仪	协同完成 0/1 级数据定义文档以及相应的 toml 文件；根据 0 级数据定义生成相应仿真图像；加入独立 QC0 数据处理流水线；开展定标参考文件对接定标流水线的工作。
6	积分视场光谱仪	协同完成 0/1 级数据定义文档以及相应的 toml 文件；启用并完成基于 Jenkins 平台的新版仿真软件的单元测试；利用 RSS 流水线对新一组硬件测试数据进行了分析。
7	星冕仪	协同完成 0/1 级数据定义内容完善及相关文档撰写工作；完成星冕仪近红外相机的性能测试及非破坏性读出需求的提出；完善星冕仪观测编排程序优化，尝试在编排中考虑背景光和杂散光的影响；进行观测仿真程序 2.0 版本开发。
8	太赫兹	协同完成 0/1 级数据定义更新工作，获得了更新版的 HSTD M 0/1 级数据的定义文档、demo fits 以及 toml 文件，推动 1 级数据处理流水线输入输出标准化程序化工作；结合 HSTD M 天津测试，初步规划 QC0, QC1 mask。
9	天体测量	确定了单帧图像相对定位星表的数据内容及解释，形成文档；完成了天体测量一级数据产品的 demo fits 一级 toml 文件；完成了定标数据文件天体测量部分的定标数据说明的补充。
10	观测数据仿真软件	持续开展新版仿真软件的开发测试任务，改正星表导入接口定义的问题，升级 PSF 数据矩阵并更新 PSF 和像场畸变的生成工具，升级平场、暗场等参考图像的仿真接口和数据命名规则，升级头文件模块的定义接口，完成多组仿真图像和参考图像的测试工作；对仿真软件的各功能模块开展单元测试工作，更新仿真软件的使用手册，组织讨论了新版数据的仿真工作和数据说明书的更新工作，以满足新版软件发布的各项需求。
11	数据可用性标记	协助完成 1 级数据定义文档编写及会签；进行各模块 QC 的检测列表初步汇总。
12	在轨定标	定标参考数据定义逐步收敛到比较完备的版本；定标产品工具包的讨论和确定，巡天模块已上传 PTC 工具包，进行定标产品 Cycle9 单元测试的讨论；随着 0 级数据定义文档更新巡天模块 bias/dark/ledflat/gain 的 toml 文件；进行超级天光平场和测光平场新方法的调研及预计开展的算法升级；测试使用 GU 和 GI 波段灯谱仿真数据拟合分析；MCI 在轨定标产品新增并制定了 6 种参考文件的格式；MCI 的 OSBS, OSDK 和 OSFT 分别合并到 BAIS, DARK 和 FLAT 文件中；IFS 单次曝光的平场参考文件类型名确定为 STDSTAR。
13	观测需求编排	讨论单元测试方案并梳理测试项，为单元测试整理第三方库依赖包，实现网站启动脚本，通过另一脚本来启动网站在线测试，补充一般观测申请的记录型操作的单元测试代码；优化编排比较工具前端；进行 Aladin 前端界面定制、动态天区变化等功能开发；开展警报监控端优化设计。
14	数据处理软件集成	编写数据处理软件的数据流图；推进 0/1 级数据定义，提供 toml 与 fits 验证的代码支持；推动流水线运行界面数据视图的讨论和开发；推动主巡天仿真数据定制工作；基本完成紫金山实验室数据迁移；配合测试团队进行仿真和编排进行单元测试，提供测试平台。

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 3 月	完成 0 级数据定义更新版本会签（已完成）	2024 年 7 月	完成 C9 流水线集成测试
2024 年 3 月	完成 1 级数据定义会签（已完成）	2024 年 7 月	完成编排软件单元测试
2024 年 4 月	完成定标参考数据定义会签	2024 年 7 月	完成配置项软件设计评审
2024 年 5 月	完成仿真软件 3.0 版本发布	2024 年 7 月	启动 1000 平方度仿真
2024 年 6 月	完成 C9 流水线单元测试和集成封装		