空间站光学巡天科学应用系统总体方案

工作组成员：陈建生 黄茂海 江林华 黎建辉 刘怡 孟宪民 陕欢源   
王锋 武剑锋 吴开超 许优华 苑海波 詹虎 张天萌

目 录

[1 项目背景 3](#_Toc528572616)

[2 任务目标和总体要求 3](#_Toc528572617)

[3 系统构成与架构 3](#_Toc528572618)

[4 总体方案 5](#_Toc528572619)

[4.1 数据产品与要求 5](#_Toc528572620)

[4.1.1 数据产品 5](#_Toc528572621)

[4.1.2 数据产品要求 6](#_Toc528572622)

[4.2 巡天仿真分系统 6](#_Toc528572623)

[4.2.1 图像仿真 6](#_Toc528572624)

[4.2.2 巡天编排仿真 13](#_Toc528572625)

[4.3 数据流水线分系统（各位） 17](#_Toc528572626)

[4.3.1 多色成像图像处理流程 17](#_Toc528572627)

[4.3.2 空间站望远镜无缝光谱数据流水线 23](#_Toc528572628)

[4.3.3 无缝光谱波长和流量定标 24](#_Toc528572629)

[4.4 数据质量保证分系统 25](#_Toc528572630)

[4.4.1 必要性分析 25](#_Toc528572631)

[4.4.2 数据质量保证工作机制 26](#_Toc528572632)

[4.4.3 影响数据产品质量的环节分析 26](#_Toc528572633)

[4.4.4 数据质量保证措施 26](#_Toc528572634)

[4.4.5 主要功能 29](#_Toc528572635)

[4.4.6 系统组成 29](#_Toc528572636)

[4.5 数据服务分系统 31](#_Toc528572637)

[4.6 观测运行支持分系统 31](#_Toc528572638)

[4.6.1 必要性分析 31](#_Toc528572639)

[4.6.2 主要功能 32](#_Toc528572640)

[4.6.3 系统组成 33](#_Toc528572641)

[4.6.4 关键问题 35](#_Toc528572642)

[4.7 软件研发支持分系统 35](#_Toc528572643)

[4.8 数据设施分系统 36](#_Toc528572644)

[4.8.1 数据中心 36](#_Toc528572645)

[4.8.2 服务中心 37](#_Toc528572646)

[4.9 工作模式与信息流程 37](#_Toc528572647)

[4.10 接口关系 37](#_Toc528572648)

[5 关键技术与解决途径 37](#_Toc528572649)

[6 研发组织与管理 37](#_Toc528572650)

[7 集成测试方案 37](#_Toc528572651)

[8 研制流程与计划 37](#_Toc528572652)

[9 人力与经费需求 37](#_Toc528572653)

[10 现有研制基础与保障条件 37](#_Toc528572654)

[11 研制保障条件需求 37](#_Toc528572655)

[12 小结 37](#_Toc528572656)

# 项目背景

空间站光学巡天是我国载人航天工程的重大科学项目，将由2m口径的巡天号光学舱实施。这个巡天立足于2020-30年代国际天文学研究的战略前沿，在科学上具有极强的竞争力，将与欧美同期的大型天文项目并驾齐驱，优势互补，并在若干方向上有所超越，有望取得对宇宙认知的重大突破。

由欧洲Euclid（1.2m口径，日地L2点）、美国WFIRST（2.4m口径，日地L2点）和LSST（8.4m口径，地基）项目的重要成员以及其他知名专家组成的国际咨询委员会对项目的科学意义予以了高度评价，同时也指出了数据管理系统（大致对应科学应用系统）、科学队伍和科学准备工作投入严重不足的风险。欧美类似的空间项目在发射前投入模拟仿真、数据管理系统研发和预先研究等工作的人力达到了上千人年（Euclid：750人年，GAIA：1400人年），历时10年左右。美国在哈勃空间望远镜（HST）发射的9年前，专门成立了空间望远镜科学研究所（STScI），负责HST的科学运行，其中一项重要工作就是研制、运行和维护HST的科学应用系统。

科学应用系统对于数据采集后的科学研究不可或缺，同时对望远镜硬件的研制也起到重要的支撑作用，所以应当与硬件同步开展工作。考虑到光学舱研制工作已进入方案阶段末期，预计将于2024年前后发射，但其科学应用系统的工作尚未启动，因此时间已经极其紧迫。研制科学应用系统需要投入大量人力，亟需启动相关工作，以切实保障空间站光学巡天能够充分发挥其潜力，取得重大科学成果。

# 主要任务与功能

光学舱科学应用系统的设计依据是《光学舱科学要求》和《光学舱科学应用系统要求》，前者明确科学目标和可定量考核的任务要求（软硬件及运行的顶层要求），后者进一步明确从科学运行到科学产出的链路的顶层要求。其主要任务与功能如下：

* 定义满足科研需求的数据产品，制定数据质量要求，建立若干典型应用场景和大型课题的应用场景（运行仿真、图像模拟、数据处理、数据分析等）；
* 根据数据产品和应用场景，制定相应的数据服务的要求，数据服务内容包括数据分发、数据访问、数据可视化、数据查询与分析、大型计算分析与临时数据存储、面向科学团队的数据处理接口等；
* 定义运行仿真、图像模拟、数据处理流水线、数据服务和核心课题数据分析软件的架构与接口；
* 定义安全的底层计算、通信和存储的架构与接口；
* 评估资源需求，提供满足要求的工程解决方案；
* 研究并建立满足要求的数据处理、地面定标和在轨定标方法，开展必要的地面试验；
* 开发，运行，并持续改进科学应用系统的所有软件；
* 按照既定的方案，租用或构建并运行维护计算存储设施；
* 光学舱运行期间，支持其科学运行，处理数据，并提供数据服务。

# 系统构成

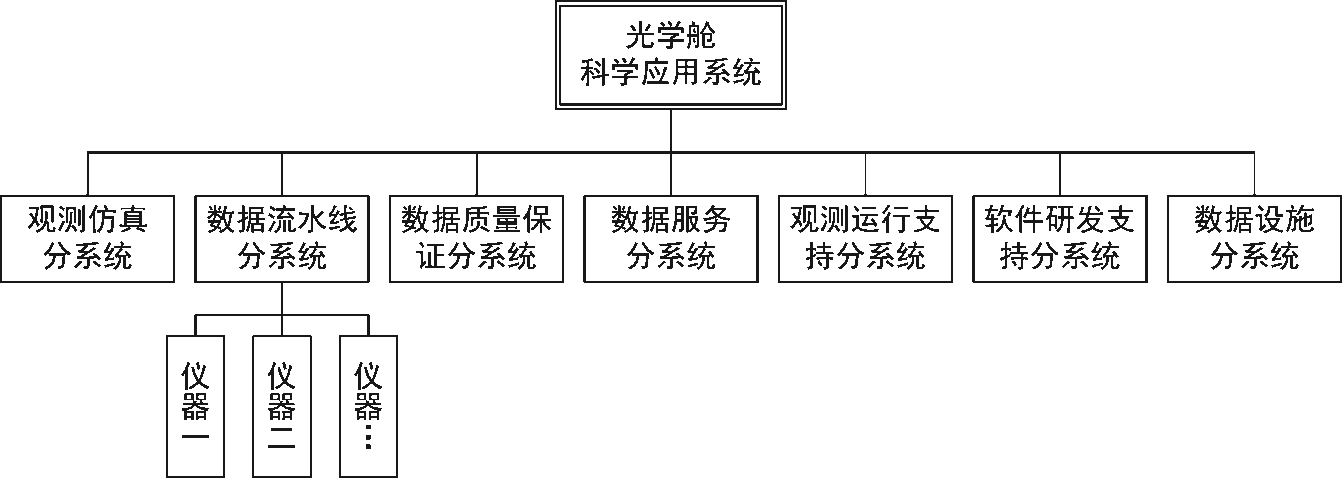


图 5 光学舱科学应用系统的构成

如图 5所示，光学舱科学应用系统包括以下7个分系统：

观测仿真分系统

进行光学舱的运行仿真和模拟观测，生成模拟数据，用于研究硬件方案和运行条件对观测结果的影响，调试数据流水线及其下游的其他分系统和数据分析软件。

数据流水线分系统

科学应用系统的核心，将原始图像加工成满足不同研究方向共性需求的数据产品。

数据质量保证分系统

研制过程中的科学质量控制和运行过程中原始图像和数据产品的科学质量监控。

数据服务分系统

通过前端的用户图形界面和后端的高性能数据库，提供高性能数据查询、数据分发、部分数据分析等服务。

观测运行支持分系统

负责光学舱的科学运行，如进行巡天运行的编排，处理其他天文模块的观测申请，规划入选的观测等。

软件研发支持分系统

制定科学应用系统的软件架构，定义接口、协议、规范等，全面支持科学应用系统的软件研发和科学家的数据处理与数据分析软件的研发。

数据设施分系统

存储光学舱观测数据和提供计算能力的硬件设施。科学应用系统的软件系统不应过度依赖于硬件架构，两者的研制应分离，避免软硬件捆绑。可考虑有偿使用国内已有的大型计算存储设施。

上述分系统中，数据流水线是核心，需要投入大量天文学家开展研发工作。虽然一些现成的天文软件可以将数据处理到看似可用的水平，但不能满足下一代大型巡天项目对数据产品的质量要求，因此各大项目都研发专用的数据处理流水线。光学舱同样需要研发自己的数据处理流水线，而且数据处理流水线的改进将一直延续到巡天完成之后多年。

# 总体方案

## 数据产品与要求

### 数据产品

空间站光学巡天数据产品分三大类型：

原始数据：包括全部需要保留的成像数据、光谱图像数据和工程数据，以及相应资料，例如观测计划、报告、定标文件等；

科学图像和光谱数据：经过校准并扣除各种仪器系统误差后的单次曝光图像、叠加图像(NUV/u/g/r/i/z/y波段)以及一维光谱数据。包括相关附加信息，如观测信息(曝光时间、观测质量、观测对象的定标文件历史等)，点扩展函数模型、光学畸变图像、屏蔽区域（mask）等；

星表数据：天体的物理性质参数列表，包括测光和光谱数据以及经数据处理后得到的物理参数及其误差 (如位置、流量、星等、颜色、分类、星系形状、测光和光谱红移等信息)。

每级数据产品不仅包括处理生成的数据，还包括与之相关的质量控制数据，从而保障所有数据以及其处理过程的可追溯性。

另有其他两类非常有用的数据产品：

外部数据：其他空间和地面巡天项目数据；可用于流量定标、数据校准(经过空间站望远镜数据处理流程再分析，校准星系形状以及测光红移测量)等；

模拟数据：全过程数值仿真；通过模拟真实数据校准和验证数据处理流程以及观测策略。

### 数据产品要求[斟酌]

空间站光学巡天以精确测量暗能量属性为驱动性科学目标，其数据产品须满足以下要求 (参考Euclid卫星数据产品要求，具体数值需结合全过程数值仿真进行估计)：

巡天区域内用于弱引力透镜研究的星系平均密度：ng≥30gals/arcmin2；

星系形状测量精度：乘法偏差m<2×10-3, 附加偏差方差σ2sys<10-7；

背景星系红移分布对应的中位数红移：<z> > 0.8；

红移0.2<z<2.0范围内，测光红移误差的均方根(RMS)： σ(z)/(1+z)<0.05；

测光红移灾难性失效比例：fcat<10％；

每个红移区间的红移分布n(z)的平均值精度：σ(<z>)/(1+z)<0.002。

## 巡天仿真分系统

### 图像仿真

针对科学需求的前期数据模拟和仿真是项目预研过程的一个重要环节，它从硬件方案和运行条件出发，生成高度可信的观测图像，以科学为闭环，优化软硬件和运行方案，并为数据处理系统的研究提供基础。

在数据模拟和仿真的范围内，将会依次开展的几个方面的工作有：

构建现实化的模拟星表。

构建模拟星表的工作是在望远镜正式运行之前，理论和观测之间的首次结合，是进行图像仿真等工作的基础。因此，各大项目的科学家都投入很大精力去制作模拟星表，如Gaia (Slezak & Mignard 2007), Pan-STARRS1 (Cai et al. 2009), LSST (Gorecki et al. 2014), Euclid (Ascaso, Mei & Benitez 2015) 等。

构建模拟星表的要求和内容有：

模拟星表的构建应尽可能接近实际观测，并达到必需的深度。实测星表如HUDF、CANDELS、COSMOS等，其观测设备的分辨率和深度都比较高，可以做为参考星表。CANDELS项目使用HST的WFC3和ACS成像设备，GOODS-N、GOODS-S、UDS、EGS、COSMOS场的观测数据， F814W波段极限星等达到27.72等，随着其星表数据的不断发表，将促进广泛领域下的天文研究，尤其在河外科学领域。COSMOS (Laigle et al. 2016) 星表包含了HST的空间观测以及地面望远镜的观测数据，提供了约118万个河内及河外目标的数据，覆盖天区为2平方度，探测深度在i波段为26.2星等。Deep Lens Survey是一个地基观测项目，其探测深度可达到B、V、R、z' 波段29、29、29、28星等每平方角秒，巡天面积为7个2°×2°的天区（Wittman et al. 2002）。此外还有很多河外巡天项目，都是可利用的重要的参考星表。

基于数值模拟的星表也是重要的研究内容，可以与基于实测的星表进行对照和互补，例如Millennium Simulation Observatory项目构建出的模拟星系表（Overzier et al. 2013）。

波段覆盖空间站光学巡天的所有测光波段，即NUV, u, g, r, i, z；同时尽量向更大的波段范围扩展，以便和Euclid、WFIRST、LSST等项目进行数据融合与互补。

星表的参数中除了天球坐标、各波段星等和误差以外，还要提供星系的尺度、星系类型、形态等信息，用于计算星系的信噪比、面亮度密度，用于样本选源的参数。

根据科学需求在星表中考虑强、弱引力透镜模型，将用于构建包含引力透镜的仿真图像。[[应该是一个自洽的星表]]

构建一个发射线星系的模拟星表。按照现有的发射线星系表，将各发射线的流量加入到各波段的模拟的亮度数据中，测试发射线的引入对测光红移的计算造成的影响，并研究如何改进其测光红移的计算。[[6、7以及成像的样本应该是同一个样本的子集]]

构建一个用于无缝光谱图像模拟的星表，深度要高于无缝光谱的探测极限点源GI~23.9等（深度巡天），做为无缝光谱图像模拟的参考星表。

计算参考星表在空间站光学巡天各波段上的星等。通过SED拟合与测光红移的计算，得到每一个天体的连续谱模型；再对连续谱模型卷积滤光片透过率曲线，计算天体在各波段内的理论流量值；在泊松分布的模型下，随机产生模拟的流量值及其星等，从而得到各个天体的模拟数据。

科学仿真图像的生成

图像仿真在天文学研究中已经发挥越来越重要的作用。通过图像仿真可以为准备研制的天文观测设备评估提供重要的依据，仿真的图像可以用于数据处理方法和流程的研究、制定和编写。当设备正式工作后，可以以最快的速度完成数据处理方法的验证、修改，并尽早投入使用。天文研究依据仿真图像可以估计各观测量的测量精度，评估对科学理论做出的限制，因此随着天文观测领域的蓬勃发展，各个大型项目如LSST、Euclid、WFIRST等都要求图像仿真能够满足其对于高精度图像处理和分析的日益苛刻的需求。天文研究对图像仿真提出的需求主要有：

生成具有广泛代表性的各类天体的成像模型，包含模型轮廓、多波段的颜色效应、成像系统的PSF模型并模拟其在视场内的分布等；

考虑到天体间的混叠对数据处理造成的影响，包括星系-星系之间、星系-恒星之间以及恒星之间的混叠等，考虑星系面亮度轮廓与光学成像PSF的卷积；

引力透镜的研究需要在图像模拟中引入引力透镜模型，对天体进行剪切、放大、旋转等几何变换，并满足引力透镜研究所需要的高精度；

要模拟探测器像素噪声、像素内的响应不均匀性、天光背景、宇宙线等；

各类天体的无缝光谱模型、系统的光谱相应函数及其卷积结果；

仿真过程可调节，运行速度快，以便通过与实测数据对比，调整算法、星表和参数。

成像观测的图像仿真主要有两种实现途经。一种是根据模拟星表，基于各类天体的模型构建星系、背景及噪声等，并最终合成图像；另一种是基于多波段实测图像，按照模拟星表中各波段的星等，经过一系列处理，通过波段的内插产生空间站光学巡天的仿真图像。[[大天区？]]

基于模型的构建方式的主要实现方案为：

PSF模拟：因为光学舱是空间观测设备，不存在大气扰动带来的星象弥散，所以PSF仅来自于观测设备的特性；

星系和恒星的模型：依据星表中星系光度、尺度、倾角、形态分类等的分布信息，构建各类星系的模型。星系按面亮度轮廓模型进行构建，为其增加积分流量、尺度、扁率、倾角方向等信息；依据星表中的恒星信息，构建PSF，以一定的数密度生成恒星的模型图像；

背景及噪声：噪声主要包括读出噪声和信号噪声，信号噪声在主要包含天光背景噪声和来自星系的信号的所产生的随机噪声、像元间相关噪声等。

基于实测图像的构建方式：

对星系形态更为精确的模拟需要使用真实的观测图像，特别是当需要模拟的星系为相互作用星系、并合系统，以及星系内部有多处恒星形成区的情况。空间站光学巡天可用的天基观测的真实星系样本主要来自HST的观测。由于真实星系图像中含有各种噪声及像元缺陷，因此将真实样本用于图像仿真要经过如下一些过程：

获取基于实测图像所产生的星表，得到其中各个天体的各波段星等、红移、SED等信息；

计算各星系、恒星在各波段内的亮度；

对原始真实图像去除像元间相关噪声（实现方法可参考LSST项目的Galsim用户手册）

将各天体图像按亮度生成各波段的图像（此时图像中带有原始图像及处理过程中的各种噪声）；

对HST的真实星系图像进行Shapelet建模；Shapelet建模的效果可参考图5-2-2；

对Shapelet模型做（HST）PSF退卷积；

在需要时，对退卷积的真实图像做引力透镜的剪切、会聚变换；

与空间站光学巡天的PSF卷积；

将所得图像在探测器像素空间做采样和插值计算，以随机的旋转角度和位置生成天区仿真图像；

估计所在天区的天光背景，将其加入仿真图像；

将各类噪声加入图像中，包含星系的光子噪声、各类随机噪声及像元间相关噪声；

以上两种实现途经都将展开相应的工作，通过对比两种方式所产生的图像，发现各自的优缺点，然后根据工作需要选择具体的实现方式。

除恒星和星系之外，在图像仿真中要考虑到的各种要素有：

引力透镜效应；

考虑探测极限以下的天体对背景的影响；

太阳系天体，如各大行星、小行星带等；在做好巡天运行仿真的基础上，可以计算各太阳系天体在某天区内出现的时间；

黄道光、地气光、杂散光及本星系群、银河系等对图像背景的贡献；

天体在成像过程中的采样问题；

快门效应；因为快门的打开和关闭时间较长，所以对快门造成的曝光不均匀性要进行改正；

宇宙线；

河外天体的银河系消光；

各像元的量子效率不均匀性；热点/列、坏像元/列；

望远镜镜片、改正镜和探测器的反射、折射、衍射过程对像质的影响；

电荷在探测器中的横向扩散的影响；

饱和像元及溢出；

探测器效应，如：读出噪声、暗电流、放大噪声、本底、电荷转移效率等；

焦面布局对各波段像质的影响；

望远镜跟踪精度；

光学系统和探测器元件的表面平整度、准直性。

图5-2-3展示了光学系统设计、望远镜跟踪、衍射、探测器和镜片的微形变、探测器点扩散函数等因素对LSST的PSF的影响效果。

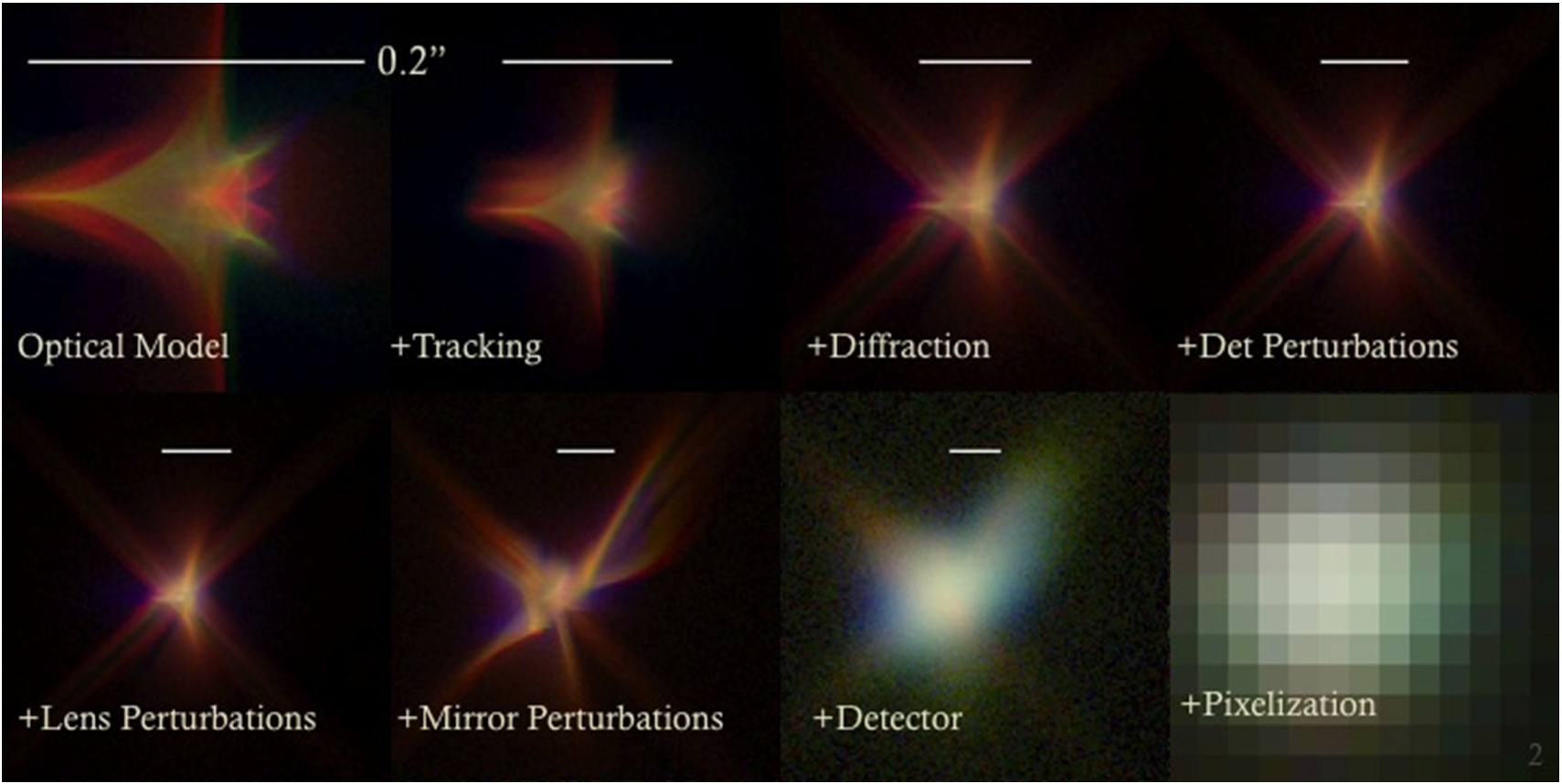


图 1各种仪器效应对LSST PSF的影响

无缝光谱图像的模拟

首先，参照无缝光谱的点源极限星等（GI<23.9）制作模拟星表；

依据模拟星表中天体的坐标安排天体色散的参考点的位置；

采用合适的天体SED模板或真实的天体光谱，在一定的波段内（GU、GV、GI），按照无缝光谱分辨率R~250对SED/光谱进行采样；

采用发射线星系表，在SED上发射线波长处增加发射线信息；

点源色散采样后的光谱宽度为PSF的宽度；

对于面源，将其划分成若干个点源的空间集合，再将其中每一个点源进行色散模拟；

在探测器像素空间进行像素化处理；

模拟天光背景、信号噪声及仪器噪声，最终生成无缝光谱的仿真图像。

现有仿真结果

现有仿真主要基于HUDF和COSMOS生成的的模拟星表，采用星系的Sérsic模型进行图像仿真。首先，采用Coe, Benitez等人发表的哈勃超深场（HUDF）星表，得到星系的位置、亮度等信息，该星表同时提供了星系的Sérsic轮廓指数；结合Beckwith等人（2004）发表的提供了FWHM、有效半径和椭率的HUDF星表，构建了星系的模型；利用APASS的星表进行恒星图像的模拟，最终生成了模拟的HUDF图像。

图5-2-4展示了*g，r，i*三个波段和这三个波段叠加之后的模拟图像。从图中可以看出，在*ｉ*波段可探测的星系更多一些。

图5-2-5展示了GV波段、6000×4000像元尺寸的无缝光谱仿真图像。采用了zCOSMOS的真实光谱进行色散模拟、zCOSMOS样本的星系形态参考了HUDF，使两个样本的形态参数的分布一致。

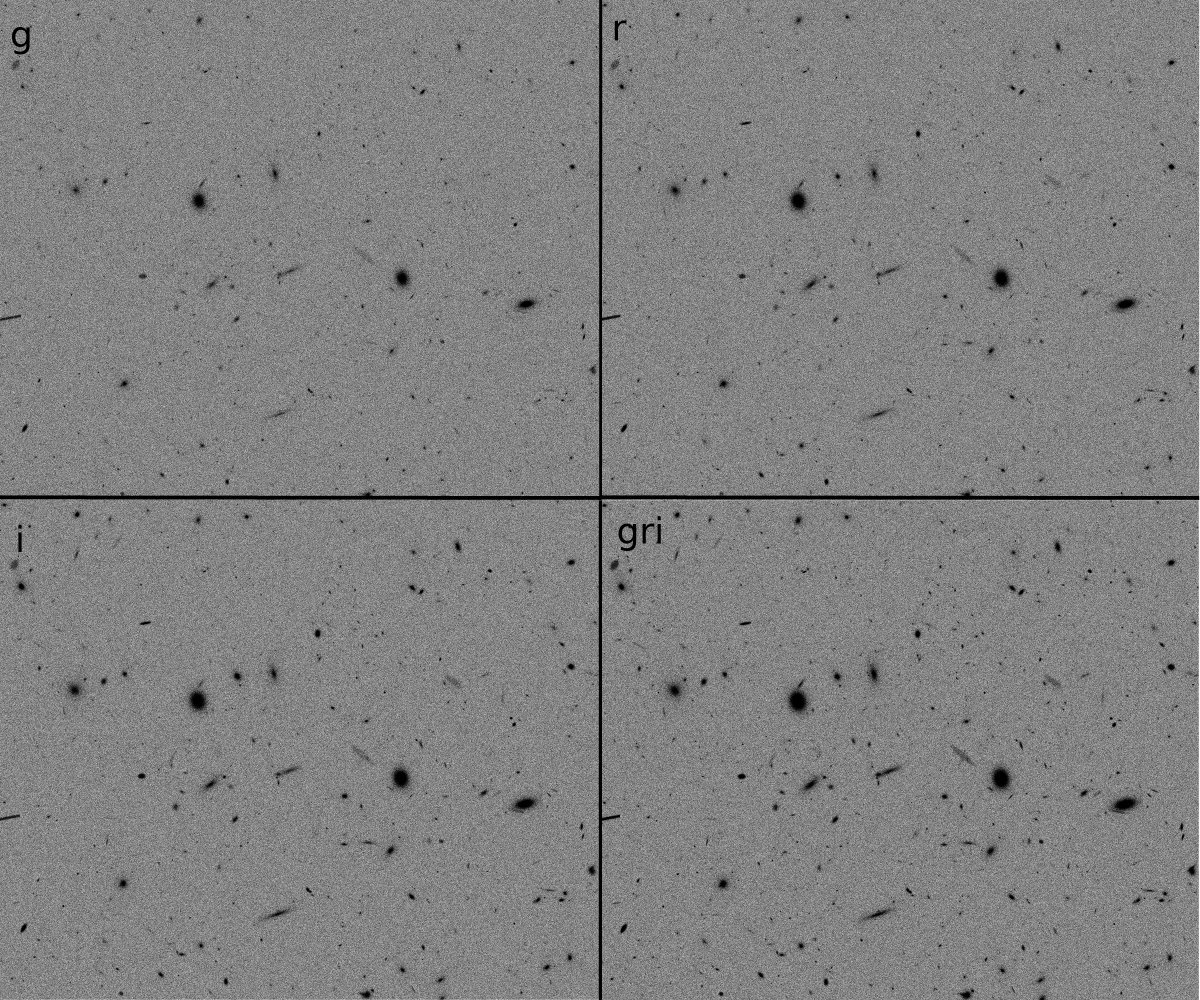


图 2 使用HUDF星表图像模拟结果

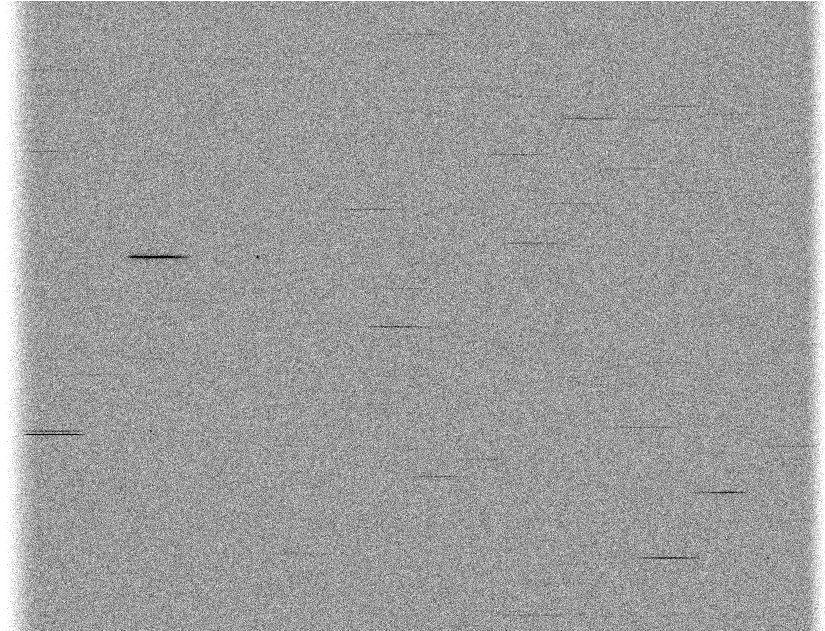


图 3 无缝光谱仿真图像（400-650nm）

### 巡天运行仿真

重要性和意义

巡天运行仿真可以比较真实地模拟出在规划的运行时期内的巡天任务完成情况，对优化飞行器平台、光学设施和巡天观测终端的硬件方案都有十分重要的作用。巡天运行仿真生成的天区覆盖、天文观测条件等信息，可加入后续的图像模拟工作中，以便生成更加符合实际观测过程的图像样本，尽可能真实的模拟包含巡天观测、数据处理、科学分析的完整过程。这对于进一步深入论证和精确量化科学产出是不可或缺的。

巡天运行仿真中积累的资料和经验，也会为后续实际的观测运行提供基础。例如，可以在现有编排仿真程序的基础之上，开发出功能更加完善的高效率的巡天观测指令生成器，能够在保障核心巡天任务能够完成的前提下，灵活地规划观测时间用于对一些持续时间短的具有重大意义的天文现象进行观测。

仿真的内容

巡天运行仿真的主要研究内容是在给定的约束条件和可观测天区搜索策略下，比较真实地模拟望远镜在轨时寻找可观测的目标天区、转向目标天区和曝光的过程，论证空间站巡天项目的科学目标的可行性。仿真结果以时间序列的方式呈现，其中包含了在某一时刻的望远镜观测指向、天文观测条件、曝光时间和转向下一个观测天区的机动方式等信息，可供其他环节的研究和论证使用。

巡天运行仿真中所考虑的约束条件包括了诸多不利于天文观测的影响因素和巡天模块自身硬件条件的限制。目前所考虑的各项约束条件主要包含以下几项：

太阳与望远镜视轴的夹角不得小于50度，月球与望远镜视轴的夹角不得小于40度；

望远镜视轴与地球亮边的夹角不得小于70度，与地球暗边的夹角不得小于30度；

望远镜飞过南大西洋异常区（SAA）上空时不进行观测曝光；

太阳能帆板的法线与太阳和望远镜之间连线的夹角须保持在正负25度之间，确保运行期间整个巡天设施有充足的能源供应；

在一轨时间内，望远镜的转动角度以及相应的机动次数必须满足控制力矩陀螺（CMG）的限制要求，不得过于频繁地进行大角度的机动。

以上1、2、3项约束条件为静态的几何限制条件，可根据当时的望远镜指向、太阳、月球以及望远镜所处的轨道位置直接计算出来，在巡天运行仿真程序中可以很方便地实现。然而，第4、5项约束条件则是随着时间动态变化的，并且受到过去一段时间内运行状态的限制，需要实时计算。此外，与硬件有关的约束条件还包括CCD的尺寸、焦面布局、相邻观测区域间的重叠大小等，这些直接影响了天区划分的方式，并决定了完成既定巡天面积任务所需的最少曝光次数。

已有的仿真结果

目前的巡天运行仿真结果表明，17500平方度的深度多色巡天和400平方度的极深度多色巡天目标覆盖面积在10年的运行时间内是可以完成的，能够满足既定科学目标的需求。图5-2-7展示了巡天运行10年后的天区覆盖情况；图5-2-8展示了巡天面积随运行时间的增长关系。

未来需要进一步开展的研究工作

目前的巡天运行仿真工作仍有许多可以改进的地方，需要继续投入精力进行研究和完善。我们将主要针对以下方面进行研究：

进一步细化各项约束条件的物理模型，深入全面地研究各项约束条件对巡天效率的影响；

考虑亮星对巡天观测的影响；

优化巡天策略，进一步提高可观测时间的利用率；

优化观测天区的选择；

优化望远镜的机动方式，减少对CMG的消耗；

科学目标的变动对巡天任务的影响；

设计和编写用于分析巡天运行仿真结果的程序，制定与图像模拟模块的接口，数据格式等。

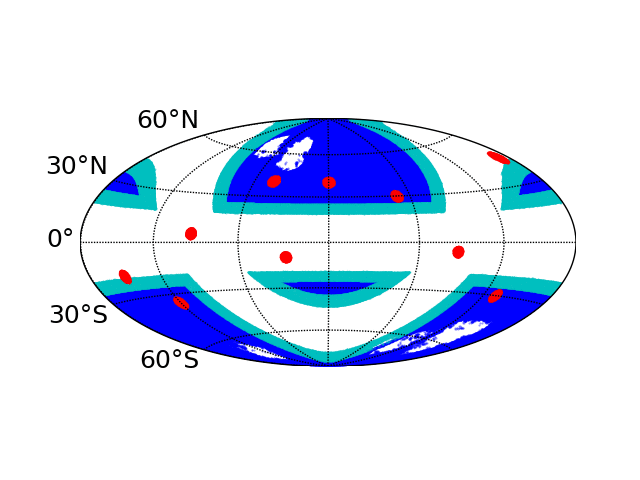


图 5 空间站巡天运行10年后的天区覆盖情况（红色标记的为极深度多色巡天区域）

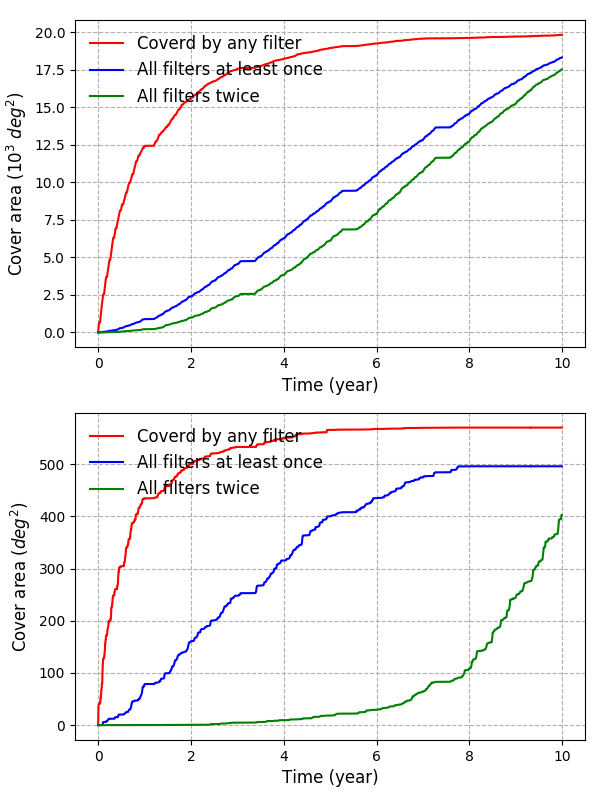


图 6 深度多色巡天（上）与极深度多色巡天（下）覆盖面积随巡天时间的增长关系。（红色表示任意波段覆盖一次，蓝色表示所有波段覆盖一次，绿色表示所有波段覆盖两次）

## 数据流水线分系统

工作包、地面定标和试验、在轨定标

### 多色成像图像处理

天文成像观测的图像处理流程，一般是指将从探测器获得的原始数字图像进行修正和定标，最终生成单图、叠加图像和残差图像，然后在图像中探测天文目标源，并进行流量和形状参数测量的过程。

#### 处理流程的输入和输出数据

完成整个流程需要的输入数据包括：

从探测器（光学上一般是指CCD相机）获得的原始数字图像（16位整型），其中包括科学观测图像，定标图像（偏场和平场）；

已释放的参考星表数据库，包括整个天空完整的已知天体位置和流量信息（例如斯隆数字化巡天和盖亚数据库等）。

经过整个流程后，输出的数据包括：

修正并定标后的科学图像，其中包括每个像元的流量、权重和mask，图像的世界坐标框架系统（World Coordinate System；WCS），流量定标参数，天光背景信息等。如有需要，还包括多次重复观测后的叠加图像或者为寻找变化天体进行图像相减后的残差图像。

在修正后的科学图像上进行的天体探测和测量，并形成包括天体位置，流量，形态，分类等信息的星表。

将图像和星表以及图像获得时候的望远镜和探测器信息进行归档分类以方便检索的数据库系统。

#### 处理流程简介

天文测光图像的处理流程可以分为以下几个基本部分：

仪器效应改正

此过程主要负责改正望远镜和探测器所造成的图像数据不均匀性。因为探测器的各个像元不可能完全一致，同样数量的光子在不同像元上产生的计数有差异，加上望远镜和探测器的不同结构也会对图像的某些部分产生影响，所以需要进行探测器修正。主要包括以下几步：

Cross-talk修正：探测器一般都是采用多片拼接技术，加上为了加快读出速度，采用多通道读出。不同通道之间在读出时候会产生不同程度的干扰，表现为某一通道的数据会在其它通道的相同位置产生计数。这个比例一般低于万分之一，但对于某些饱和星（计数大于6万），就会产生不可忽略的干扰。通过计算可以准确获得干扰系数，对图像进行修正。

Overscan修正：利用探测器边缘区域的数行或者数列，生成二维的探测器本底，并从所有图像中减去该本底。该区域不参加曝光，与图像同时采集，但因为是利用有限几行像元进行二维估计，故此需要结合本底修正一起，进行更准确的改正。

本底修正：每个观测周期前后，望远镜进行5-10次0秒曝光拍摄，通过平均或者取中值计算，获得本底图像，并从该观测周期的所有观测图像中扣除该本底图像。

线性修正：探测器的各个像元在计数值较低或者较高时候，常常存在偏离线性的情况，这对于暗弱天体的测量影响较大。通过不同曝光时间的平场图像，可以计算偏离度，并对所有图像进行修正。

平场修正：利用均匀面光源获得探测器各像元对光子的响应系数，并进行修正。地基观测中常使用圆顶平场和天光平场。每个观测周期前后，望远镜进行5-10次短时间曝光，保证每次曝光的探测器计数在1万以上，通过归一化并平均后，获得平场图像。人造平场与真实来源于无穷远的星光还是不完全一致，尤其在大视场探测器的边角部分差异较大，因此须利用科学观测图像合成超级天光平场，结合高信噪比的内部平场，可以获得准确的平场图像，并对科学图像进行修正。[根据巡天相机方案修改]

坏像源标定：探测器上的数千万个像元中，总会有少量像元不合格。通过平场图像，可以找到并在mask图像中标注该像元。在后续的测光过程中可以屏蔽这些坏像源。此外，饱和亮星会对周围的部分像元产生影响，需要通过特殊算法探测到这些饱和像元并进行标定。

条纹修正：探测器在一些波长范围可能会产生干扰条纹。通过拟合或者低通滤波，可以获得条纹的结构，并在图像中进行扣除。

鬼像修正：望远镜和探测器结构可能会在观测图像上的固定位置产生不同形状的鬼像。利用大量观测图像，可以获得鬼像的结构，并从图像中进行扣除。

宇宙线修正：探测器对高能量的宇宙线较为敏感，尤其是放置于空间的探测器，缺少地球大气层的保护，会出现大量宇宙线痕迹。传统上使用拉普拉斯算法可以有效探测宇宙线并使用周围像元流量进行拟合填充，并在mask图像中标注宇宙线。对于放置于空间的探测器，可能需要根据仪器具体情况，开发更加适合的新算法来完成宇宙线修正。

天测定标

空间站光学巡天开始运行时，预期GAIA巡天DR5数据已向全世界公开释放。GAIA DR5将包含全天10亿颗暗至21等恒星的最精确位置和自行信息。将GAIA星表与观测图像测光星表中各个天体的中心位置坐标进行匹配，计算图像的世界坐标框架系统（WCS）。完成坐标框架定标后，就可以通过WCS的关键字信息，直接将像素坐标转换到天球坐标信息。高精度的坐标框架定标可以为计算天体的准确位置以及自行提供很大的帮助。大视场探测器成像的坐标框架定标存在一些问题：一是图像畸变较大。通过提高拟合的阶数（大于4阶）可以得到较好的解决。二是电子转移效率问题（Charge Transfer Efficiency；CTE）导致探测器中间与边缘位置的星象中心产生偏差。利用参考位置和测量位置残差值迭代法得到CTE对于不同星等和位置的影响，并加以修正，最终消除CTE对于天测的影响。

流量定标

[[请苑海波整理：分为绝对定标（外部定标、标准星定标）和相对定标]]

流量定标主要有两种方法：外部定标和内部定标。主要以外部定标方法为主，内部定标进行配合和辅助。用测光软件对图像进行初步孔径测光得到仪器星等测光星表，然后与参考星表进行交叉，通过星表星等比较得到测光零点（zero point）。内部定标是利用同一批天体在不同观测图像的相互交叠区域上的星等或流量的比较来对临近区域曝光图像进行流量定标和零点矫正，从而使整个巡天天区内的流量定标区域一致。通过此方法可使定标精度比外部定标提高一个量级。巡天使用的滤光片系统与传统系统在透过率、中心波长等方面可能存在差异，会导致二者的星等差随颜色项的变化发生改变，从而导致定标零点的误差和弥散增大，因此还需要进行颜色项改正。

对目前的大视场巡天来讲，主要的流量定标方法有1、经典的标准星方法（e.g., Landolt 1992）；2、Ubercalibration方法（Padmanabhan et al. 2008）；3、Hypercalibration方法（Finkbeiner et al. 2015）；4、Stellar locus regression 方法（SLR; High et al. 2009）和5、Stellar color regression 方法（SCR； Yuan et al. 2015）。

Ubercalibration方法是目前大视场巡天中常用的方法，如SDSS、PS1、Gaia。SDSS巡天是第一个应用Ubercalibration方法的大视场巡天，但由于SDSS巡天本身在设计的时候并未考虑到Ubercalibration方法，天区与天区间的重叠并不是很充分，所以尽管在其定标的时候考虑了大气消光系数在观测夜内的变化，其在g/r/i/z波段定标精度有1%，在u波段精度只有2%。PS1巡天由于是时域巡天，同一天区有多次观测，且台址条件优于SDSS，所以PS1巡天在利用Ubercalibration方法做定标的时候尽管没有考虑大气消光系数在一个观测夜内的变化，仍然取得了约1%的定标精度。Gaia巡天由于不需要考虑大气的影响，CCD探测器又采用TDI模式，且是一个时域巡天项目，所以天生特别适合于Ubercalibration方法。Gaia/DR1数据在G波段已取得3mmag的定标精度，BP／RP波段取得了3-4mmag的定标精度，多次测量结果平均后单颗恒星的星等误差在亮端好于1mmag。预期Gaia巡天结束时其G/BP/RP 星等定标精度和误差还将有显著的提高。

Hypercalibration方法是两个独立巡天的相互定标。SLR方法利用恒星在颜色－颜色空间的分布性质，在低消光区域可以对巡天数据开展实时的颜色定标，精度可达百分之几。如果能充分考虑恒星在不同天区不同深度下金属丰度分布的变化，SLR方法的精度能进一步提升。SCR方法则充分利用大规模光谱巡天的优势，将有精确恒星大气参数的数百万颗恒星当做颜色标准星（u-g颜色精度～3%，其它波段颜色精度1-2%），可以达到mmag级别的颜色定标精度。如SDSS/Stripe82天区的颜色定标精度经过SCR方法后已经提高到了2-5个mmag，比之前最好结果好了2－3倍。

考虑到Gaia卫星提供了在当前及未来相当长时期内精度最高的测光数据，工作组成员最近正在发展一套基于Gaia数据的流量定标方法。基于Gaia数据，该方法能精确估计数十万颗DA/DC白矮星在不同波段的精确星等，精度可达2-5mmag。基于Gaia测光数据和其它巡天提供的分光数据，该方法能精确估计有分光观测的正常恒星在不同波段的精确星等，精度可达2-10mmag。

在充分考虑各方法优缺点、已有（和空间站光学巡天时期将有）测光、分光巡天数据及空间站光学巡天的成像特点（单次曝光同时获取不同天区不同波段图像，因此每幅图像都有其自身的零点和平场；背景天光由黄道光主导，暗弱(V=23.1mag, 150s曝光单个像元（0.1角秒）ADU计数分别为0.5, 5.7, 31.8, 41, 40, 16,5)；黄道光有显著梯度（黄纬15-45度间为~ 0.6 mag/30 deg = 0.02 mag/deg = 0.004 mag/ccd；同一CCD的相邻图像间有10角秒的重叠；亮端极限星等较暗，约18等）的基础上，建议以下定标策略：

[[绝对定标还是需要的]]

Ubercalibration方法：由于空间站光学巡天从策略上保证了同一CCD的相邻图像间有10角秒的重叠，在重叠区域我们预期有数十颗高信噪比的共同恒星可以用来做Ubercalibration。此时该方法的精度将主要取决于平场改正的好坏。 我们预期该方法能取得优于～5mmag的流量定标精度；

SCR方法：利用LAMOST及其它大规模光谱巡天观测过的恒星作为颜色标准星，预期能取得优于2-5mmag 的颜色定标精度；

基于Gaia的方法：利用Gaia数据得到的mmag级别的G星等及BP－RP颜色，结合LAMOST及其它大规模光谱巡天得到的恒星丰度信息，可与精确估计空间站光学巡天不同波段的星等、颜色信息，精度对NUV／U波段优于20 mmag，其它波段可达2－4 mmag，并以此为标准星，预期能达到2-3 mmag 的颜色、星等定标精度。

在具体实践中不同方法可与相互结合，相互验证，预期可以达到2-10mmag的流量定标精度，并力争1-5mmag的颜色、星等定标精度。

平场改正的精度对Ubercal方法的流量定标精度具有关键作用。为了更好的实践上述流量定标方案，我们在巡天策略和数据处理上还有如下一些建议:

选择一个密集星场进行多次移位观测，利用恒星构建大尺度平场。因为要在1度视场内都比较密集，所以应该是低银纬天区。移位观测10次即可达到32\*32的分辨本领（10＝2\*log2(32)）

选择一个10平方度左右的有丰富光谱观测数据的小天区做较短时间的曝光，使得亮端极限星等能降到15-16等左右，为利用SCR方法和基于Gaia数据的方法提供一个好的训练样本。[[短时间曝光比较困难]]

因为定标源相对望远镜来说非常亮，所以对CCD的非线性改正精度要求高；需要对brighter-fatter效应进行改正[[需要地面+在轨测试]]。

因为天光背景非常暗，所以对减偏场的精度要求非常高。

因为黄道光背景有梯度，所以在构建夜天光平场的时候需要考虑天光背景的梯度变化。

目标源探测以及流量、形状参数测量

利用已有或者根据优化算法，也可以结合已经释放的其它观测波段图像的星表，对定标后的图像进行目标源探测。对于探测到的目标源，需要进行形状测量并做出正确的恒星和星系分类。恒星和星系需要分别作出圆形和椭圆形的孔径测光。更进一步需要生成不同的点扩散函数（PSF）和星系轮廓模型，结合权重图和mask图像，进行复杂测光。如果有多幅图像叠加生成更深度的图像，需要考虑co-added测光。

大视场巡天普遍还会面对一个天光背景不均匀的问题。通过优化算法来寻找准确的天光背景梯度并进行扣除。当图像中存在较大饱和亮星或者展开面积较大的近邻星系时候，需要优秀的算法来标注这些展开的亮源，以获得准确的天光背景。

#### 对计算和存储的需求

如果不需要实时完成在线数据处理，对计算速度的要求不太严苛。但进行处理的部分步骤中，需要同时读入大量图像进行计算，对内存要求较高。此外，如果采用模型测光方法，则需要大量迭代和拟合，如果需要很快获得结果，对计算速度的要求也比较大。经过图像处理流程后，总的输出数据量大约为输入数据量的5-10倍，对存储空间的需求也较大。完成每个轮次的数据处理后，需要有专门的科学家与工程师对包括图像和星表在内的数据进行仔细检查，发现问题和缺点，督促数据处理部门进行算法改进。一般来说，大型巡天项目会每年更新数据处理流程的版本，并对已有数据和新增加的数据进行重新处理，然后进行数据释放（Data Release）。这就需要数据处理，归档和用户界面各单位的通力合作。[[I/O密集，high-throughput computing]]

### 无缝光谱图像处理

无缝光谱是空间站望远镜巡天科学项目的重要组成部分，将为巡天目标提供低分辨率光谱。无缝光谱具有很多重要的科学应用。[[无缝光谱用于构建星系红移样本，通过星系成团性、重子声波振荡、红移畸变效应限制宇宙学]]无缝光谱的数据处理与图像的数据处理相辅相成，可以归在同一流水线框架下。参照Euclid的数据流水线，对于空间站望远镜，在预处理得到Level 1的数据后(LE1)，图像处理和标定首先进行(PHOT)，得到测光信息，接下来第二步就可以进行二维光谱图像的处理和标定，并抽取无缝光谱(SPEC)。第三步是加入计算测光红移需要的外部测光以及光谱数据(EXT)。在整个过程中需要引入之前做模拟的结果进行不断的测试和验证(SIM)。将以上这些数据合并发布(MER)，称为Level 2的数据。最后通过测光数据来计算测光红移(PHZ)，通过无缝光谱来测量光谱红移(SPZ)。光谱红移可以用于验证测光红移的可靠性。这是Level 3的数据产品。[[需要更具体一点的描述]]

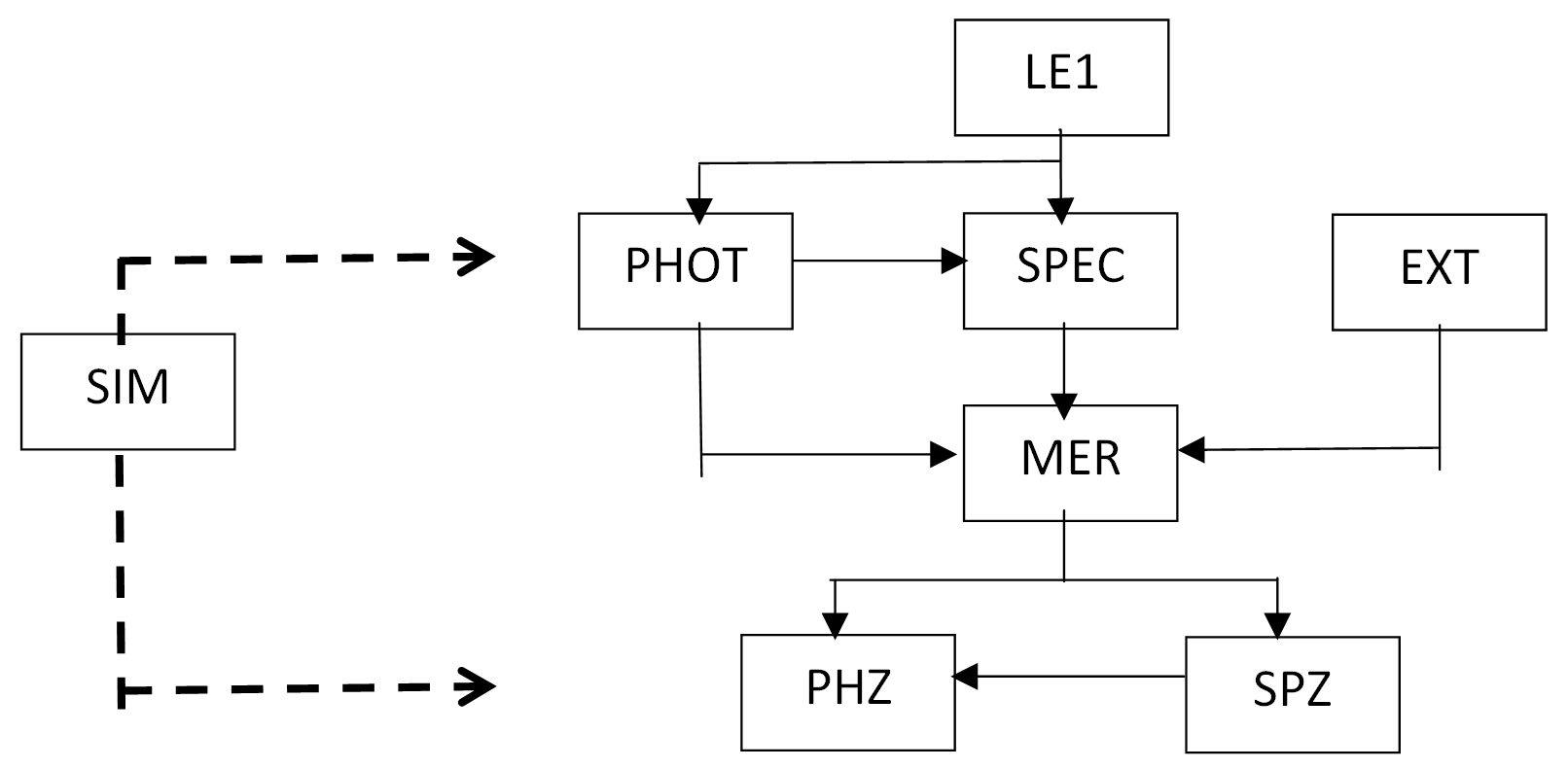


图 7 无缝光谱数据处理流程

无缝光谱的数据流水线软件可以参照HST用于处理无缝光谱的aXe Software Package(Kummel et al. 2009, PASP, 121, 59; <http://www.stsci.edu/institute/software_hardware/stsdas/axe>)。整个软件包包括三部分：

aXe：光谱抽取和定标的主程序；

aXeSIM：用来做无缝光谱二维图像的仿真；

aXe2web：用来创建可视化网页，方便用户检视aXe的输出结果。

这三部分中，前两部分是数据处理的必要组成部分，第三部分能够给科学用户带来很大方便。aXe不仅仅可以用于HST数据，也可以用于其他无缝光谱的处理。因此我们可以参照aXe Software Package来编写空间站望远镜无缝光谱的数据流水线软件。aXe本身主要用C语言编写，并融入了PyRAF和IRAF。最近（2018年5月）aXe新增了Jupyter Notebook功能。[[强调必须针对空间站光学巡天研究无缝光谱的处理方法。

无缝光谱的平场相当复杂]]

无缝光谱的定标包括波长定标和流量定标两部分：

波长定标

观测

无缝光谱的波长定标通过观测有一系列强发射线的点源来实现。例如，HST 观测的是数个行星状星云。我们同样将选择数个已经研究比较成熟的行星状星云。这些行星状星云有很强的发射线。同时，它们有非常好的地面观测数据，也就是说，它们的发射线波长是已知的（视向速度已经考虑）。为了与无缝光谱比较，地面的高（波长）分辨率光谱需要平滑，以达到无缝光谱一样的分辨率。行星状星云在空间望远镜下可能不是严格的点源，这时可以考虑别的源，例如Wolf-Rayet星。

波长基准

对于每个源，我们需要一个波长基准，也就是波长参考点。一般来说，如果无缝光谱的观测伴随着一个直接宽度成像的观测（一前一后使用同一个guide star），那么该直接成像的图像中的源位置将用作波长基准。如果没有直接成像，可以利用无缝光谱的零阶像作为波长参考点。但是零阶像往往由于色差的原因在光谱方向已经被色散开，这会给波长定标带来额外的误差。这种情况下需要额外的校准。[[用grating，0级像无色散，且像质很好]]

定标

有了上面的观测和波长参考点，我们就可以计算波长λ和距离s的关系。距离 s 指的是相对波长参考点的像元数。这儿我们假设图像已经完全处理过了，而且光谱已经扳直了。一般可以利用多项式来表示λ和s的关系：λ= a0 + a1 s + a2 s2 + …。如果简单地利用 s 作为 x 项效果不好，可以考虑较为复杂的形式。需要指出的是，这个关系一般是焦平面位置的函数，也就是说，a0和a1等是（x,y）的函数。所以需要更细致的观测和测量（见下）。

随空间位置和时间的变化

[[地面工作非常关键，在轨没有条件在焦面上进行很密集的定标]]

上面所说，波长λ和距离s的关系一般是位置（指目标源在CCD上的位置）的函数。这是由于像在焦平面不同的位置扭曲的不一样。在拍摄定标星的时候，我们需要在不同位置拍摄，以建立一个数据库。同样的，上述关系也可能是色散方向的函数（色散方向和探测器夹角），所以我们还需要在不同的色散方向拍摄定标星。

同样，波长λ和距离s的关系有可能是时间的函数。这需要我们每隔一段时间去拍摄一次定标星。目前HST 并没有测到该关系随时间的变化。

流量定标

流量定标相对波长定标简单一些。一是在空间没有大气的吸收，二是不用考虑狭缝损失流量。一般来说通过拍摄一两颗一级标准星可以把无缝光谱理论上的透过率和波长的关系计算出来。和波长定标一样，这种关系是位置和角度的函数，需要在不同的位置和角度去拍摄标准星。同样，需要每隔一段时间去拍摄一次标准星，以监测流量定标随时间的变化。

## 数据质量保证分系统

科学应用系统完成任务的质量对空间站光学巡天的顺利实施和科学产出具有重要意义。巡天期间每天数据量为~2TB，全生命周期6年获得约5PB原始数据,处理后数据总量达55PB。如何保证海量观测数据的真实性、正确性、完整性等，对于巡天科学应用系统来说是个挑战，有大量的工作要做，以保证在科学应用系统落实一系列质量策略。

因此需要在科学应用系统里专门建立一个数据质量保证系统，针对海量观测数据产品进行质量管理，支持在最短的时间内，用最小的费用建立全面的产品质量信息库、收集和分析数据产品度量数据，控制和提高数据产品质量，满足应用需求，同时减少维护费用。

### 数据质量保证工作机制

数据质量保证工作可以从质量数据采集、质量数据分析以及报告审核等多个环节进行入手，建立健全的质量保证机制和体制，规范数据处理流程，建立质量评价体系，加强质量保证过程中的监督管理等等。

### 影响数据产品质量的环节分析

影响数据产品质量主要包括几个主要环节：观测（观测数据采集）、数据流水线处理、数据流水线变更和升级。

每个环节都有可能出现问题，影响数据产品质量。例如

观测环节：设备配置和设置欠优、标定偏差、设备性能降低、空间环境影响、观测调度编排策略和执行欠优；

数据流水线处理环节：算法问题、辅助数据问题、定标问题，计算设施运行问题；

数据流水线变更和升级环节：处理软件变更问题，流水线升级问题，在规模数据处理设施部署问题。

### 数据质量保证措施[[精简一下，有些质量体系的内容不太适合]]

#### 建立数据质量评判体系

建立数据质量评判标准，针对观测、数据处理、流水线变更等环节，根据数据产品特点和要求，识别质量审核的产品、规定质量审核的时机，明确质量审核方法、确定质量审核目标，并确定优先级。

优先级是动态的。随着项目的进展，可根据需要更新识别新的质量审核目标，重新确定质量审核目标的优先级。

#### 建立质量数据收集和报告机制

建立产品质量信息数据库。根据质量信息数据库，安排质量人员、进度、分析成果交流方式，形成质量保证计划，主要包括质量目标、为实现质量目标确定的质量审核项、各角色在质量审核分析中的职责、参与质量审核分析活动的时机、质量审核分析活动的内容、进度安排、结果形式等。质量人员按质量保证计划收集和保存质量度量数据。

收集到的质量度量数据按质量评价体系分析当前产品的状态，对出现的异常情况进行原因分析，分析异常情况，同时形成质量分析报告。

#### 建立和实施质量问题分析和纠正措施制度

根据质量分析报告，识别出产品的缺陷和存在的问题，对其进行客观的描述，并明确给出处理意见，包括对缺陷和问题的解决措施、解决责任人、预期完成时间。

#### 观测质量保证

观测设备和观测运行过程会导致数据产品发生质量降低，因此，关于观测的数据质量保证分为观测设备和观测运行两方面。

观测仪器设备需要优化配置，定期标定、检测或自检，尽可能追溯到法定的计量标准。定期进行设备维护。降低空间环境变化对数据的影响，使观测仪器设备的准确度和精度应满足产品的科学要求。

优化整体巡天策略，根据设备、资源、空间环境影响不断优化和调整巡天策略，保证调度策略落实。协调优化ToO观测，其它观测终端。

#### 数据流水线处理质量保证

在数据流水线处理流程中，观测采集到的低级数据经过自动化流水线处理产生高级数据产品。有可能因为算法问题、辅助数据问题、计算机平台软硬件问题导致数据产品发生质量问题，因此需要，制立产品配置计划，建立流水线软件（包含所需的辅助数据）开发库、受控库和产品库，实施软件配置管理，确保产品受控。具体措施包括：

##### 数据产品标识和可追溯性

对数据产品建立并保持产品层级明确、易于识别且具有唯一性和永久性的标识。

建立完整的产品质量记录，确保产品可以追溯到生成过程中每一个模块的相关信息。

##### 数据产品流水线配置管理

数据产品由数据处理流水线生产。生产过程中，数据处理流水线的配置要纳入配置管理。每一个模块以及所用到的辅助数据都要建立标识和版本标识。

##### 计算设施运行

由于巡天计算量大，假设规模处理流水线运行在大量服务器上，那么相关计算和网络资源的配置、优化、检修、升级等工作将影响到数据处理的质量，所以需要对流水线运行实施以产品质量保证为目标的管理。

#### 数据流水线变更和升级质量保证

##### 流水线软件三库管理

由于流水线会在全周期发生演化和升级，流水线配置管理采取“产品库”、“受控库”、“开发库”三库管理。

开发库每次构建（build）后立即更新，包含最新的流水线软件变更或缺陷修理，软件单元经过单元测试和自动化集成测试、回归测试；

受控库是流水线在正式投入使用前的“release candidate”，每月（TBC）更新，包含相对稳定的流水线，对软件熟悉的高级用户和相关缺陷提出者进行试用并发现、报告缺陷；流水线质量检验人员通过处理规模数据鉴定库中流水线是否符合进入产品库。

产品库存有正式上线的、规模生产产品的流水线软件，相对稳定，文档齐全，每半年到一年（TBC）更新一次。

##### 开发环境和工具管理

算法研制过程中、流水线软件开发中，必须使用正版开发工具，计算机专机专用，定期使用最新版本的正版杀毒软件进行病毒检查和清楚病毒。

##### 规范流水线部署流程

位于产品库的流水线软件在例行部署到规模运行的数据中心的过程中要遵循相应的计划和预案，做到升级、生产两不误。

### 主要功能

分系统的主要功能是完成或支撑落实数据质量保证措施，包括：

支持科学应用系统的数据质量评判标准的定义、各个环节数据质量保证要求的确立和工作内容的制定。

收集质量数据，进行评估分析，生成数据质量报告

支持质量问题解决

监督观测质量保证工作的落实

监督数据流水线处理质量保证工作的落实

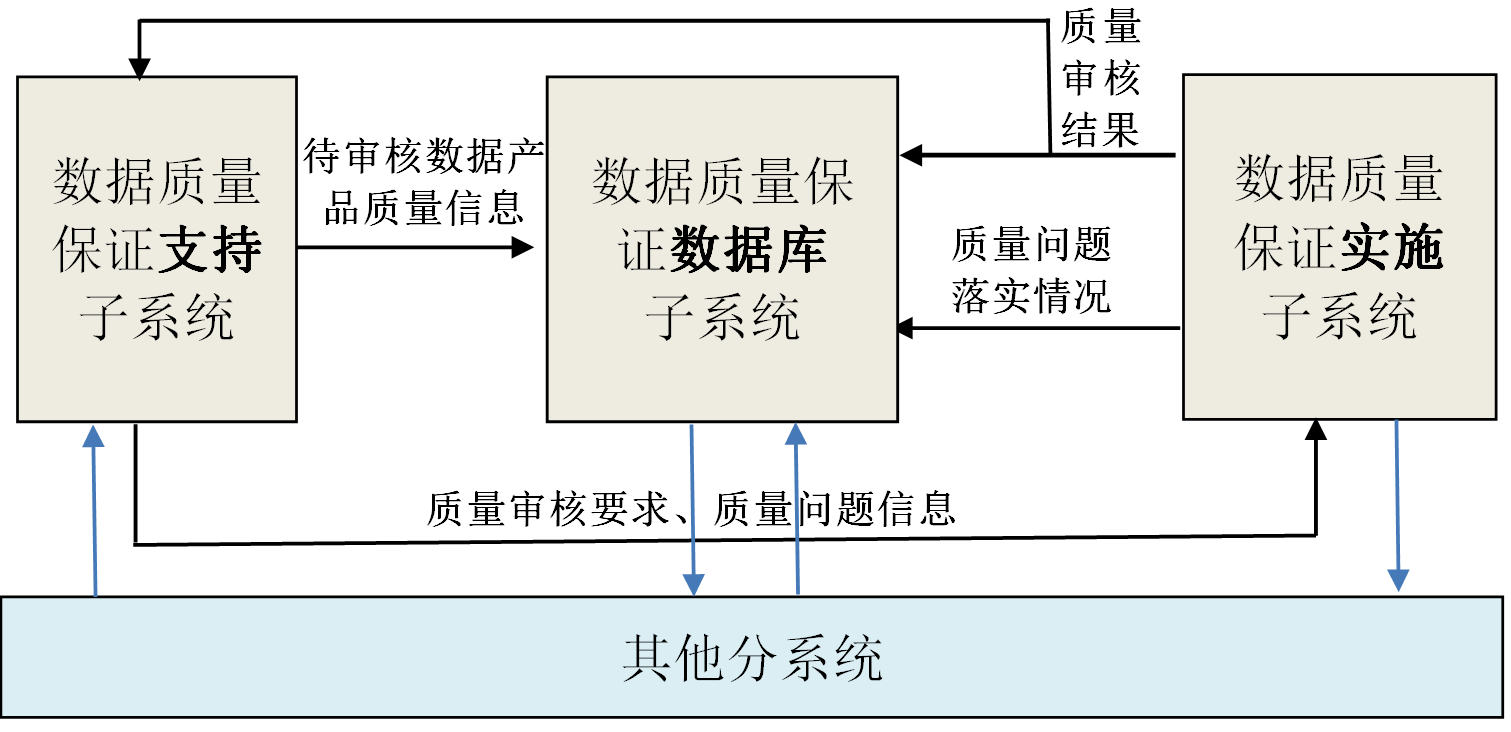
监督数据流水线变更和升级质量保证工作的落实

### 分系统组成

本分系统分为如下子系统：

数据质量保证支持子系统、数据质量保证数据库子系统和数据质量保证实施子系统。

1. 数据质量保证支持子系统：负责定义、修改、更新和发布数据产品生成过程质量要求和规范（相关负责人评审通过），组织相关人员共同定义需要进行质量审核的数据产品和质量审核的要求，负责组织相关人员共同识别质量问题和处理意见等工作；
2. 数据质量保证数据库子系统：负责建立、录入、更新和维护数据产品质量信息数据库，发布和通报数据产品质量信息等工作；
3. 数据质量保证实施子系统：负责定义和更新质量数据度量计划，根据度量计划收集、分析和判定质量数据，辅助解决质量问题，监督质量问题的落实情况等工作。



## 数据服务分系统

[[通过前端的用户图形界面和后端的高性能数据库，提供高性能数据查询与分析、数据分发、数据可视化、大型计算分析与临时数据存储、面向科学团队的数据处理接口等。需要一些具体的方案描述，不仅是功能描述]]



数据服务分系统依托公共的数据存储和公共数据库[[为什么是公共的？]]，提供产品处理服务、产品分发服务、产品收集服务、数据管理服务、数据发布服务。由数据流水线分系统负责数据文件实体和元数据信息的流转和调度，数据产品质量保证分系统进行产品质量检验并生成产品质量报告。

[[数据流水线分系统负责处理数据，生成数据产品]]产品处理服务：负责数据产品的标定、拼接、去重等处理，并生成各级数据产品；

产品分发服务：负责将各级数据产品根据分发策略向科研团队或科研人员进行实时分发；

产品收集服务：负责将科研人员处理的高级数据产品进行集中汇聚；

数据管理服务：负责进行各级数据产品统一归档与管理，包括产品的在线、离线处理、订阅申请处理等；

数据发布服务：面向科研人员和公众提供数据统一发布服务门户，提供数据快速检索获取服务，提供数据订阅服务等。

## 观测运行支持分系统

科学应用系统的核心任务为准备上行数据；对下行数据进行分析生成科学产品；分析、监控巡天模块的健康状态，在必要时进行定标与维护操作。

观测运行支持分系统将支持其中的上行数据的准备和星上载荷的健康状态监控。将提供一系列工具，辅助科学家将天文观测计划，根据观测目标、观测约束、卫星及巡天模块工作状态，生成为特定时间段内的详细仪器工作计划，提交给上层管控中心进一步形成上行指令，由巡天模块执行；在观测执行过程中，观测运行支持分系统根据下传的科学数据和工程数据进行巡天模块工作趋势分析，为巡天模块的定标、维护提供依据，在必要时制定定标维护计划。根据观测任务的执行结果，对观测质量进行分析，为随后的观测计划编排提供依据；在分析过程中，发现可能的仪器工作异常，并及时进行排查和维护。

因此观测运行支持分系统对于巡天模块的在轨运行是必不可少的。

### 主要功能

本分系统的主要为在科学目标的指导下，在工程要求和运行要求约束下配合上级运管中心，完成观测提案收集与管理、观测计划制定、巡天模块科学性能分析、巡天模块状态及观测状态分析与演示三大类功能，从而支持巡天相机的在轨测试和在轨运行。

本分系统的基本功能要求如下图所示：



图 8 观测运行支持分系统功能要求

主要功能需求包括：

观测提案收集与管理，负责发布观测提案征集指南、提供工具支持观测提案编写、收集、管理、帮助和反馈等。

巡天任务的编排和优化仿真，输入为卫星平台状态、轨道状态、载荷状态、观测任务列表、初始任务编排计划等，对输入的观测任务列表根据各种约束条件进行执行仿真，返回对平台和载荷观测模式的影响，对观测任务编排进行优化。

常规观测计划制定，输入为常规观测提案，其中包含对于观测仪器、观测条件等约束信息，结合卫星轨道计划等信息，输出为具有特定格式的详细观测计划，包括详细载荷及卫星平台配置、观测时长、建议起止时间等。观测计划将作为输入，进一步由上级管控中心细化上行命令序列。

机遇目标观测计划制定，功能要求与常规观测计划类似，但是机遇目标的时间约束比较严格，需要在短时间内完成计划制定。

载荷故障恢复计划的制定，载荷异常的情况下，由仪器团队制定仪器故障恢复策略，形成载荷故障恢复计划。

完成载荷性能趋势分析，在一定周期内对载荷的科学观测情况进行分析，掌握载荷的工作情况，对载荷的常规定标、附加定标计划的制定提供依据。

观测计划演示，作为观测任务编排、观测计划制定的附加功能，以图、表、动画等方式演示计划编排情况，方便科学运行团队了解整体工作计划情况

观测任务的执行状态演示，根据下传数据，分析并演示当前观测任务的执行状态。

载荷工作性能，根据下传数据，分析并演示当前载荷的工作状态。

快视数据分析与演示

### 分系统组成

观测运行支持分系统的组成及其与特定任务模块配合数据流如下图所示：



图 9 观测运行支持分系统组成

观测运行支持分系统包含如上图所示的6个子系统：

观测提案收集与管理子系统：负责提供工具辅助观测提案申请人完成观测提案的编写、提交、申请反馈、观测状态反馈等工作；辅助提案遴选工作；

观测任务编排与优化仿真子系统，将通过遴选的观测提案，根据观测任务的类别、卫星及载荷状态、已有观测任务等约束条件，完成一定时间内观测任务的编排和优化。

观测计划制定与演示子系统，根据优化后的观测任务进一步根据载荷状态、轨道等约束形成将完成长、中、短期的常规观测计划（包括常规定标计划），需要紧急处理的机遇目标观测计划和载荷故障恢复计划，在本地暂存或等待提交任务中心生成观测命令序列上传。同时提供一定时期内观测计划模拟执行演示。

观测任务执行与演示子系统，将根据星上下传的数据，分析载荷观测任务的实时执行状态并予以演示；将根据观测数据分析结果，对观测任务的执行质量进行分析，形成分析报告；将提供历史任务执行情况回放演示等，辅助用户全面了解巡天相机的观测任务执行计划、现状及历史。

载荷工作性能趋势分析与演示子系统完成载荷性能的趋势分析并以，将以图、表等形式演示。

快视数据分析与演示，对于某些需要快视的数据，提供分析与演示功能，辅助科学家及时进行决策（如ToO策略、仪器特征数据等）。

### 关键问题

当前阶段巡天项目运行支持分系统面临的需要解决关键问题如下。

任务计划：将巡天仿真分 系统的结果应用到实际任务编排中并给予结果反馈。需要界定与巡天仿真分系统的界面接口。

多设备协调观测：解决巡天和其它使用望远镜设备的资源共享问题（多通道光谱仪；IFU；光星冕仪）。

故障树：为了有效预判和诊断巡天运行的故障和问题，需要建立系统（主要包括在轨设备）的故障树。

观测质量模型：建立评价体系定量评估科学运行质量。

## 软件研发支持分系统

科学应用系统所有软件研发和用户数据分析软件接入的支持（架构、规范、接口、协议、文档、调试、集成、测试、通讯等）

建立公共的软件研发支撑环境，支持以下的三种软件研发及数据分析的模式：

用户下载数据或通过接口调用数据，在本地完成应用开发（编码、调试、测试）、分析处理、结果可视化；

用户下载采样数据或通过接口调用数据，在本地完成应用开发，并按统一规范要求打包成应用模块，将本地用户自定义应用模块及相关工具上传到软件研发支撑环境中做分析处理，并在云中进行结果可视化等；

用户通过在线开发环境的支持，利用数据访问接口调用数据，通过基于工作流引擎组合用户自定义模块、标准模块形成完整的处理流程，完成直接在云中完成应用开发、分析处理、结果可视化；

科学应用系统中的软件研发和用户数据分析软件接入都可以通过以上三种模式来支持实现。

公共软件研发支撑平台主要应包括以下功能：

规范体系：主要包括管理规范、技术规范等。定义了巡天科学系统的元数据、数据规范，包括数据定义、组织等；应用系统软件模块接口规范；

数据访问接口：面向外部数据访问的、基于RESTFul的数据对象访问接口；各个软件模块（用户模块、基本工具集）之间的访问接口；

基础工具集：针对数据处理的标准化的基础模块及工具，包括基础算法、通用处理、可视化等

工作流引擎：通过工作流来组合基础工具集、用户定义软件模块等软件模块；

天文大数据管理平台（软件）：数据分布及分析处理，为上层各类科学应用提供有效支持；支持容器化的应用部署、面向应用的数据分布、归一化名字空间的EB级存储；

web化的软件开发及集成环境：用户定制软件模块上传及管理、工作流定制及管理，支持用户应用管理、相关子系统的软件研发；软件开发的代码、调试、测试；软件开发过程管理（版本管理、bug追踪、持续集成等）

## 数据设施分系统

### 数据中心

数据基础设施基于分布式的多数据中心建设，通过400Gbps以上的高速网络互联，支持大数据管理平台、公共软件研发支撑环境、各个业务系统以及各类用户科研应用的运行。

数据基础设施按分层结构考虑，可分为三个层次：

项目自建部分：这部分是基础及核心，支持项目的基础运行、技术研发，部署基础系统软件、核心数据等的基础环境；

基于国家科研设施部分：基于国家、中国科学院信息化基础设施来建设，主要功能包括：数据在线存储、数据分析处理等；

基于公有云部分：可利用公有云的能力实现外部扩展，主要功能包括：数据分发、容灾备份、国际数据交换及互联互通等。

数据基础设施能力可以分为存储能力、计算能力，基于分层结构的分布式数据中心实现。每个分布式数据中心内部，又包括若干个服务器集群，支持数据存储、分析处理。

从硬件设备类型看，分为网络设备、服务器设备等。按存储能力、计算能力的需求，结合数据特点、应用特点，定制不同配置的外存、内存、CPU的各类服务器设备；网络设备则实现服务器设备互联以及网络访问、安全管理的功能。

### 服务中心

面向科研人员以及相关管理人员，为项目运作提供7×24的全天候的服务支持、培训服务。内容覆盖专业基础知识、数据分析应用、应用开发支持、专业培训等。

## 工作模式与信息流程

## 接口关系

# 关键技术与解决途径

# 研发组织与管理

# 集成测试方案

# 研制流程与计划

# 人力与经费需求

# 现有研制基础与保障条件

# 研制保障条件需求

# 小结