



广州大学
Guangzhou University

SKA成像质量评价方法研究

吴少锋 王锋 邓辉 梅盈 徐奕骏 谢扬帆

广州大学天文系/天体物理中心

云南丽江 2022.7.20



广州大学
Guangzhou University

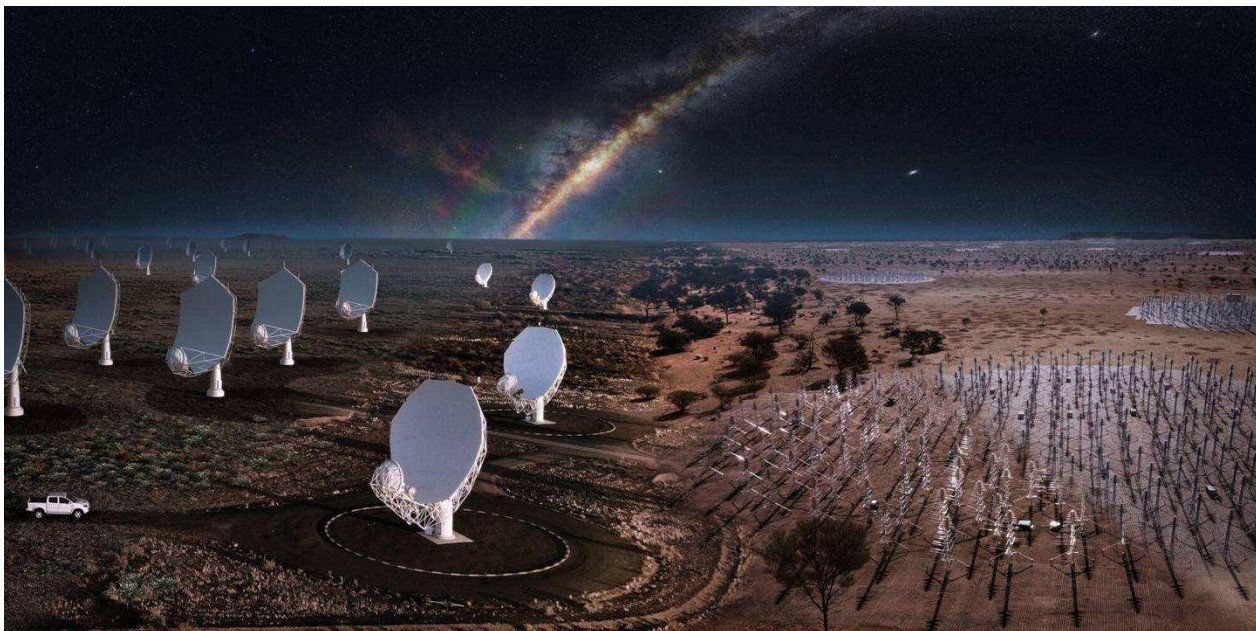
主要内容

- 研究意义与目标
- 质量评价
- 源搜寻
- 现有评价方法
- 研究计划

研究意义

- 下一代射电望远镜：平方公里阵列（SKA）

超高灵敏度、超大视场、超快巡天速度、超高频率分辨率、超高时间分辨率、超高空间分辨率



- 宇宙黎明和再电离探测
- 脉冲星搜寻、测时和引力理论检验
- 中性氢巡天和宇宙学研究
- 宇宙磁场
- 星际介质
- 暂现源探测
- AGN和黑洞
- 中性氢星系动力学和演化
- 生命摇篮
- 超高能宇宙射线低频探测

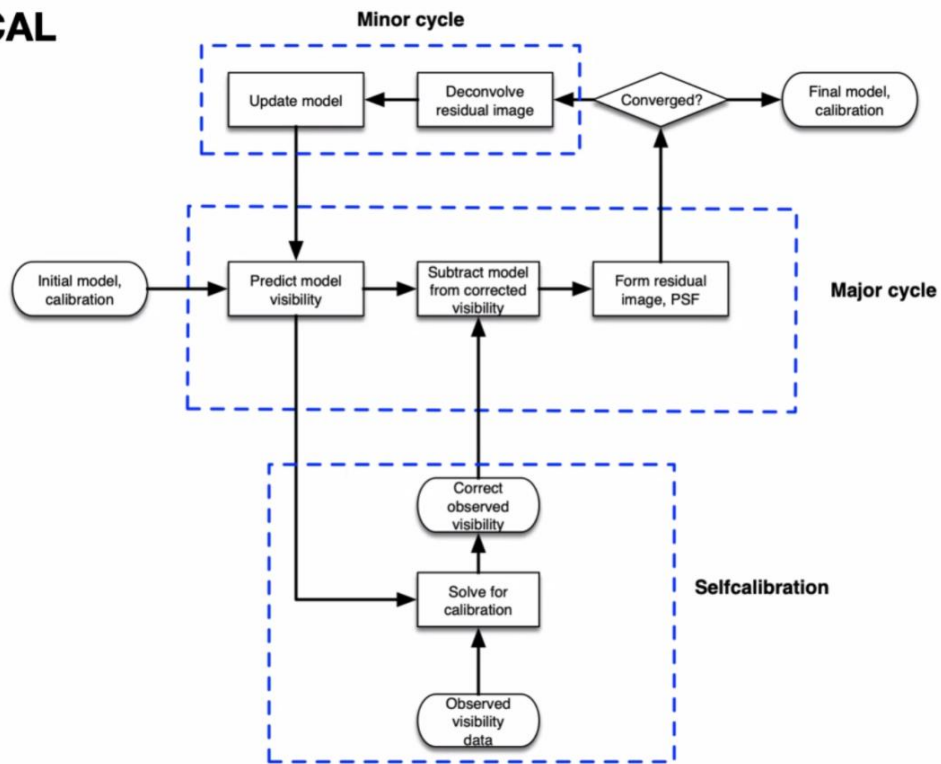
- SKA将产生前所未有的超大观测数据量，给数据处理管线带来巨大的挑战



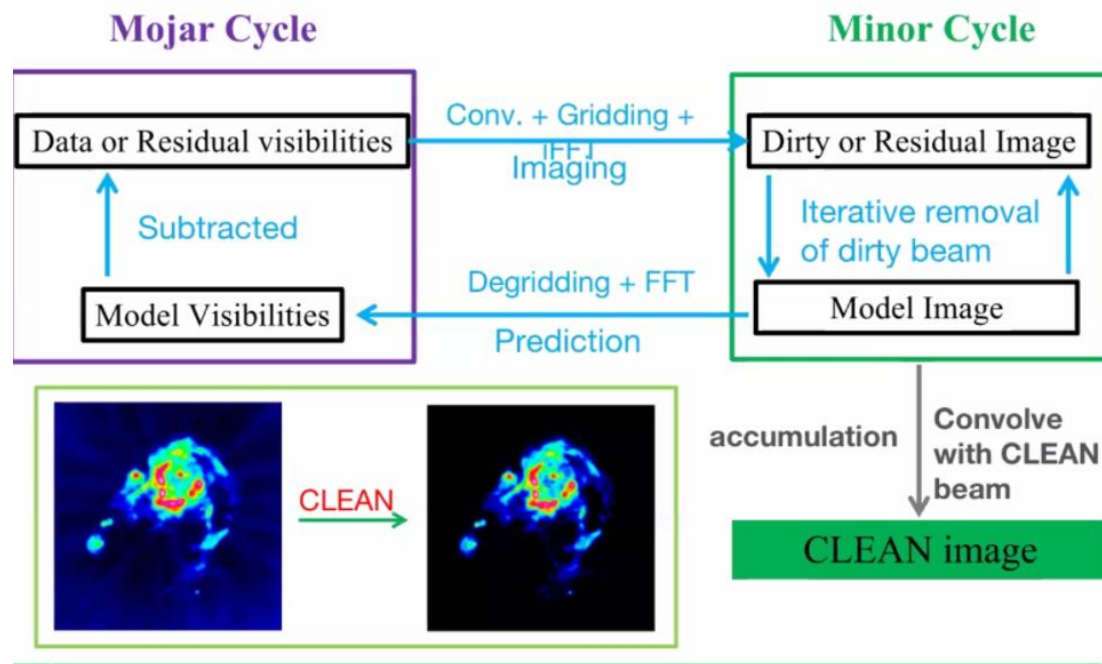
研究意义

- 数据处理的关键流程：连续成像管线——自校准和成像

ICAL



CLEAN Framework for Deconvolution



- 自校准与成像算法的科学质量与准确性 ← 质量评价

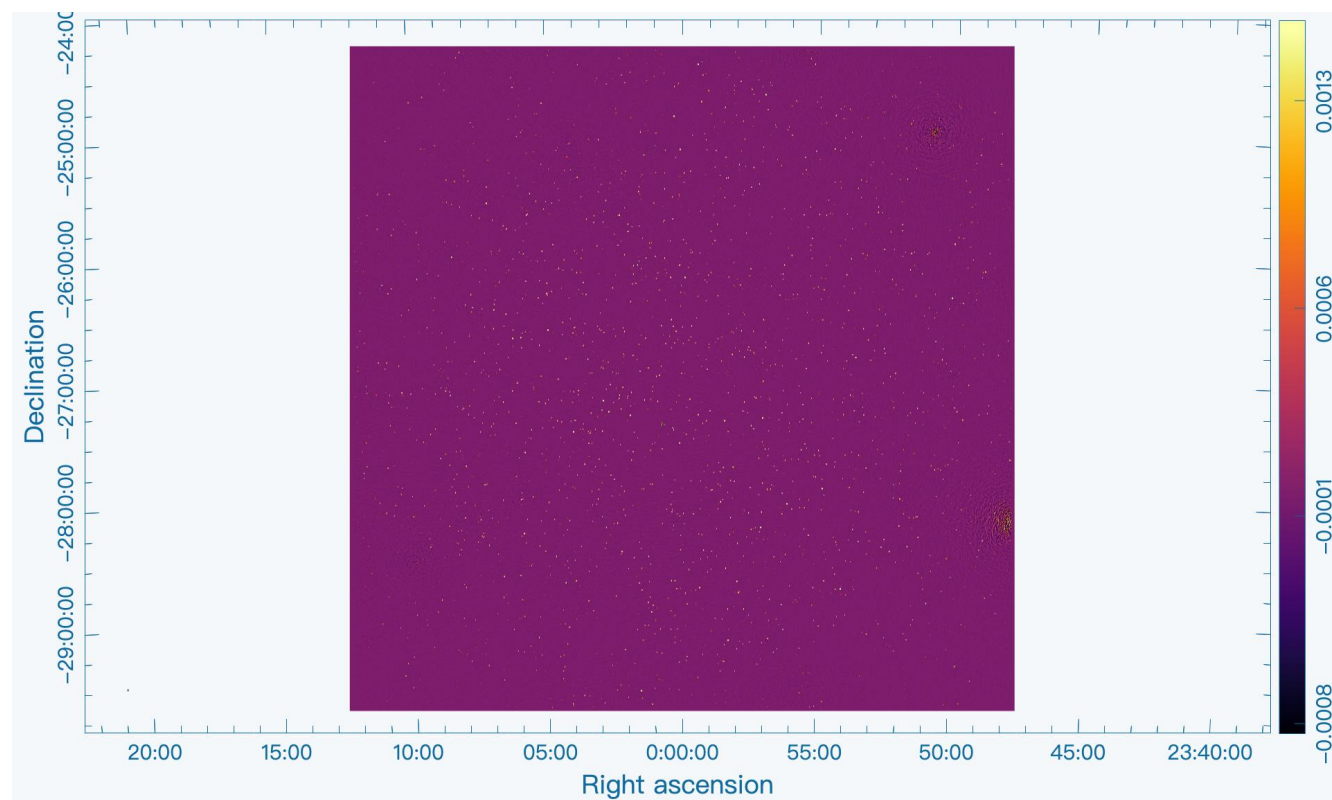


质量评价过程

- 使用已知星表进行模拟观测
- 使用连续成像管线处理观测数据得到洁图
- 对连续成像管线进行质量评价：
 1. 源搜寻
 2. 与原始星表进行比较
 3. 输出评价指标

源搜寻

- 源搜寻：寻找包含天文源信息的像素



restored图

背景估计

- 背景估计——实现背景和源分开：

通常是设定阈值， 阈值选择在尽可能检测更多的源和尽可能小的错误检测率之间权衡

- 1) 人为设定阈值 (典型值为 5σ) 2) 错误检测率 (FDR)分析设定阈值 (Hopkins et al.2002)

当图像中的背景是变化时：

Zones算法

- 将图像划分为一定数量的区域，计算各区域的背景和噪声
- bkg: 流量分布的50th percentile
- rms: $(75\text{th}-25\text{th}) / 1.349$
- 计算速度快，实施简单，但捕获不到在小于区域大小的空间上变化的背景和噪声

Grid算法

- 内部尺度 (grid) , 外部尺度 (box)
- 计算以给定的grid点为中心的box区域内像素背景和噪声
- 对区域内像素进行sigma裁剪，计算均值和标准差，屏蔽超过均值 3σ 的值，重复两次
- 比Zones慢，但更精确

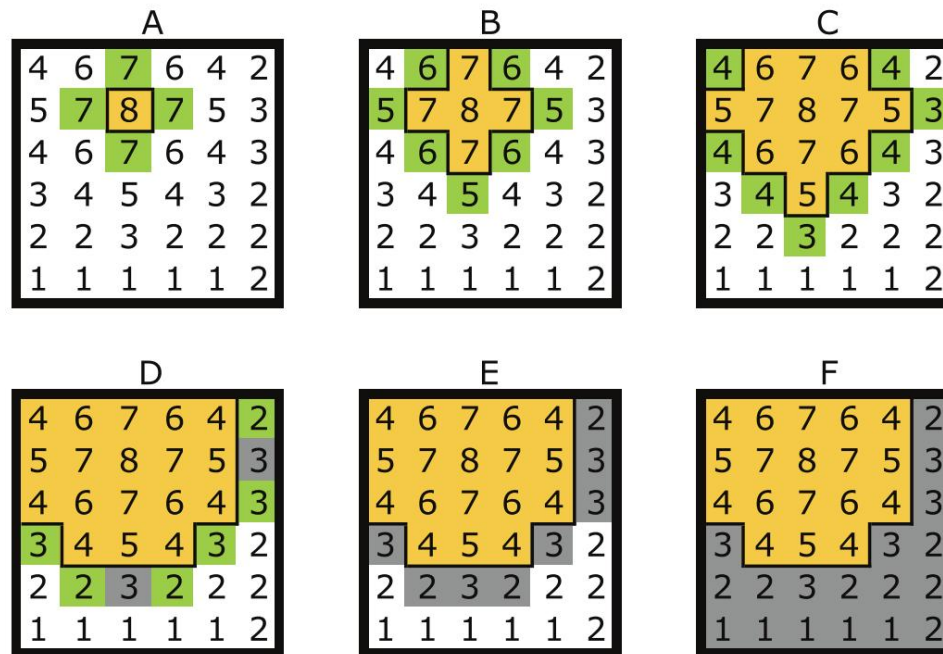
源识别

- 源识别:

将高于给定阈值的像素分组为islands, 一个island对应一个或多个源

FloodFill算法 (σ_{seed} , σ_{flood})

- 高于 σ_{seed} 的像素用来播种islands
- 高于 σ_{flood} 的像素用来生长islands



Unprocessed pixels
 Pixels in the island
 Pixels being considered
 Background pixels

源表征和形成星表

- 源表征——测量每一个源的属性：如总流量、角大小等
使用一系列高斯成分来表征island中的像素

- 多重高斯拟合：

de-blending:

将一个island分成多个子island，每个子island都适合一个成分

iterative fitting:

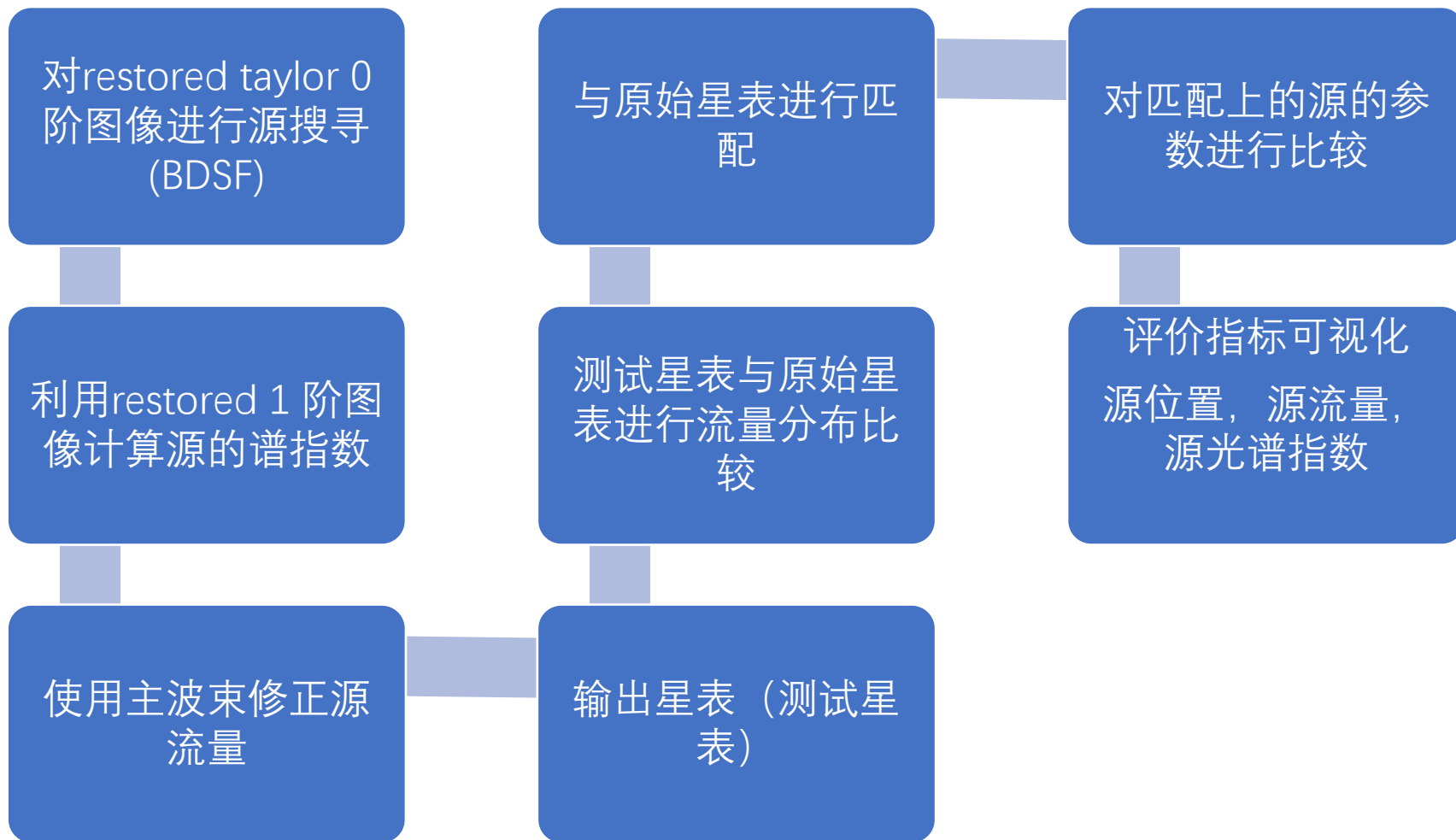
评估拟合残差是否需要额外的成分来拟合，该步骤重复进行，直到达到可接受的拟合或达到设定的成分数量

- 形成星表：

提取源的参数并形成该视场范围内的星表，包含拟合参数的不确定性和拟合失败、不佳的源



现有评价方法



retored图的0阶是指其taylor展开的第一项(MSMFS)

目前工作

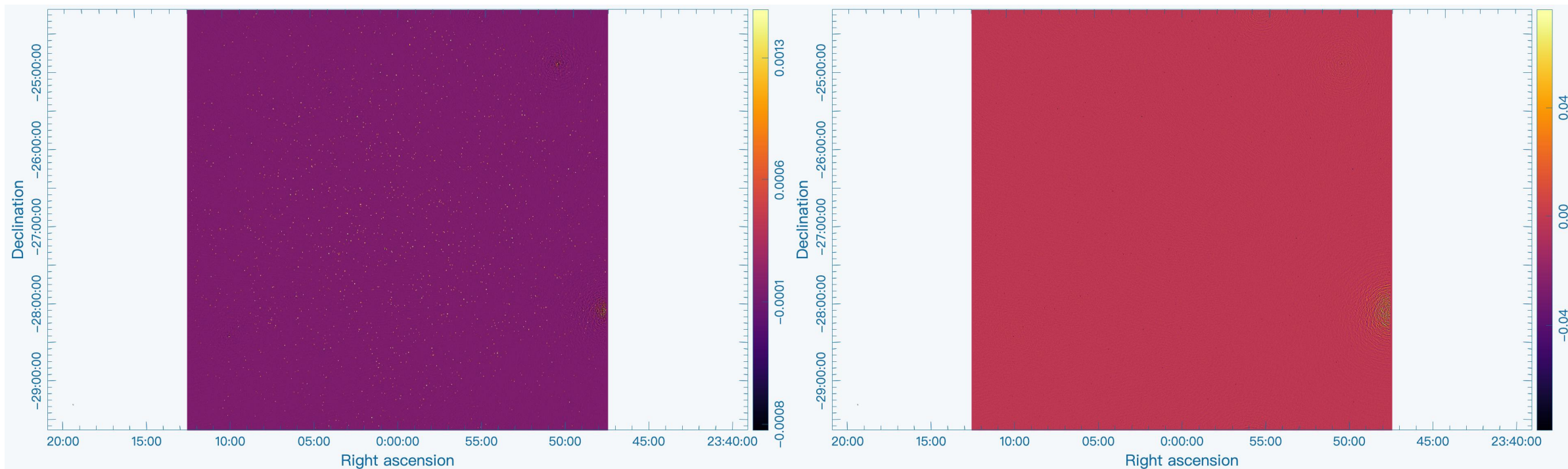
- 为了探究现有评价方法可以进行哪些改进，我们进行了模拟观测并使用了现有评价方法进行质量评价
 - 使用OSAKR进行模拟观测得到SKA1_LOW仿真观测数据
 - 使用RACIL对观测数据进行成像

Parameters	Values
skymodel-1	GLEAM
skymodel-2	random faint source (FOV < 5°)
fainter source	Minimum flux 10^{-4} Jy
telescope	SKA1-LOW full array
phase_centre	RA: 0.0°, DEC: -27.0°
start_frequency_hz	140 MHz
num_channels	30
frequency_inc_hz	200 kHz
channel_bandwidth_hz	10 kHz
observation length	14,400 s
num_time_steps	40
time_average_sec	0.9 s

Parameters	Values
clean_nmajor	10
clean_niter	10,000
clean_algorithm	MSMFS
clean_facets	8
imaging_npixel	32,768
imaging_cellsize	$2.908882e-06$ (FOV=5.5°)
imaging_weighting	robust -2.0
ingest_vis_nchan	30
ingest_chan_per_blockvis	6
clean_restored_output	taylor
clean_nmoment	2

目前工作

- 成像结果



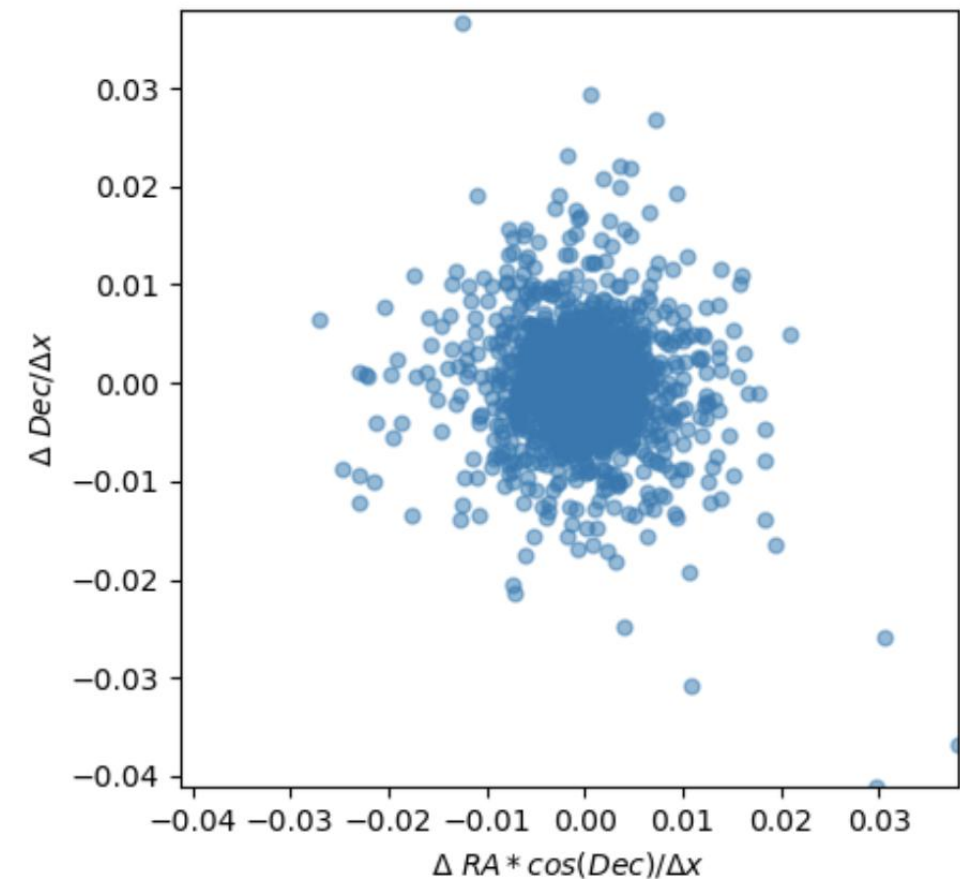
retored.taylor.0

retored.taylor.1

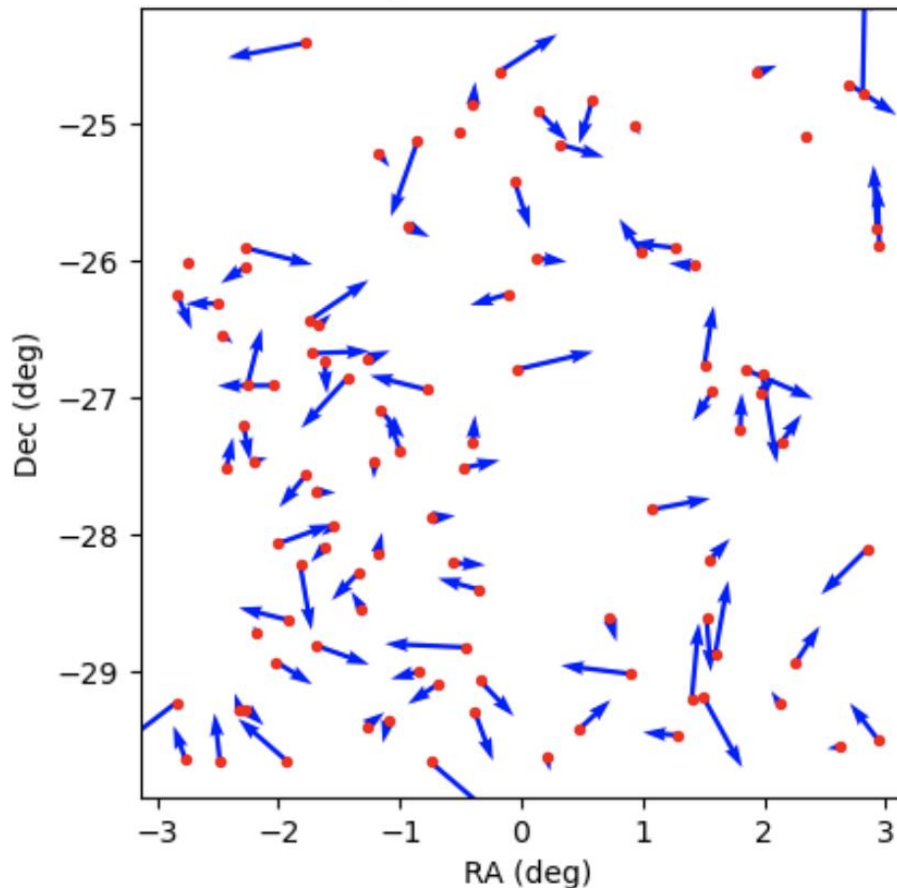
- 对成像结果使用现有方法进行质量评价

评价指标——源位置

Errors in RA and Dec



Brightest 100 sources



左图：相对于图像分辨率的匹配源之间的位置误差。

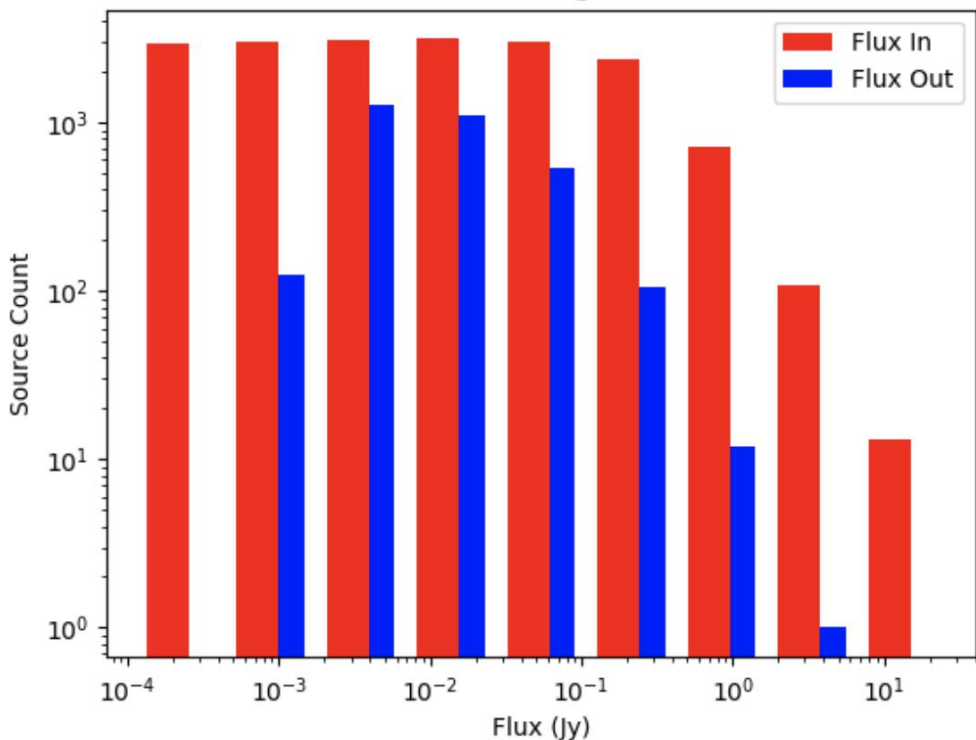
右图：最亮的 100 个源的源位置偏移图。红色代表原始星表源的位置，蓝色代表测试星表对应源的位置偏移矢量（放大后）。

- 从左图可以看出大部分匹配源的赤经和赤纬偏差在2%以内
- 从右图可以看出大部分匹配源的位置偏移是无规律的

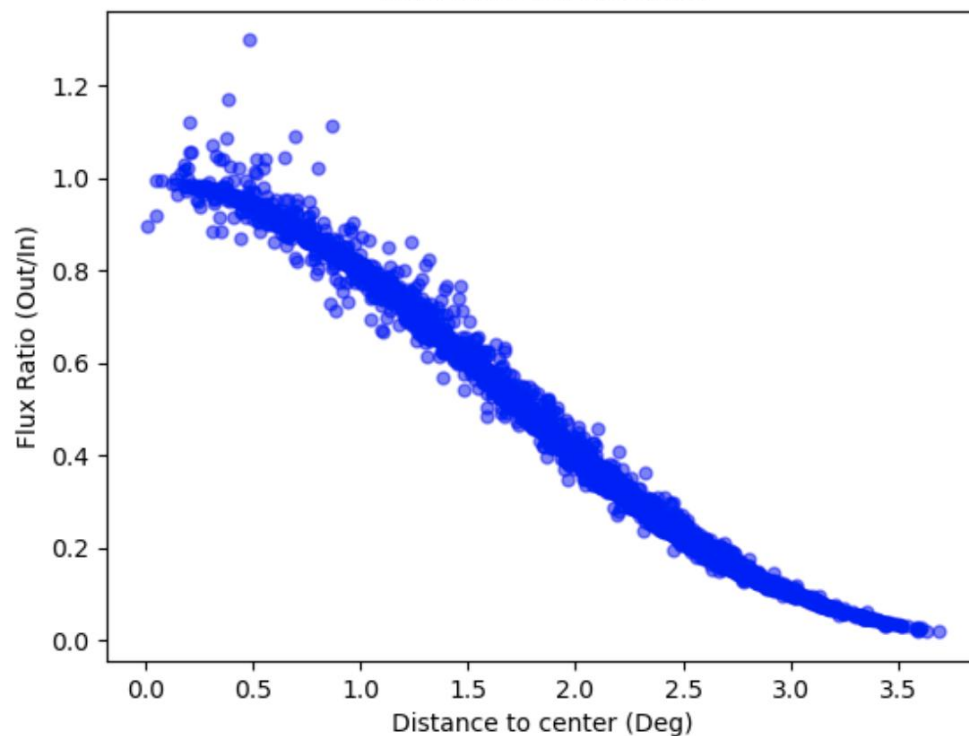
成像过程没造成位置上的系统偏移

评价指标——源流量

Flux histogram



Flux ratio vs. distance



左图：测试星表和原始星表的源流量分布图。

右图：相对于相位的距离的所有源的流量比

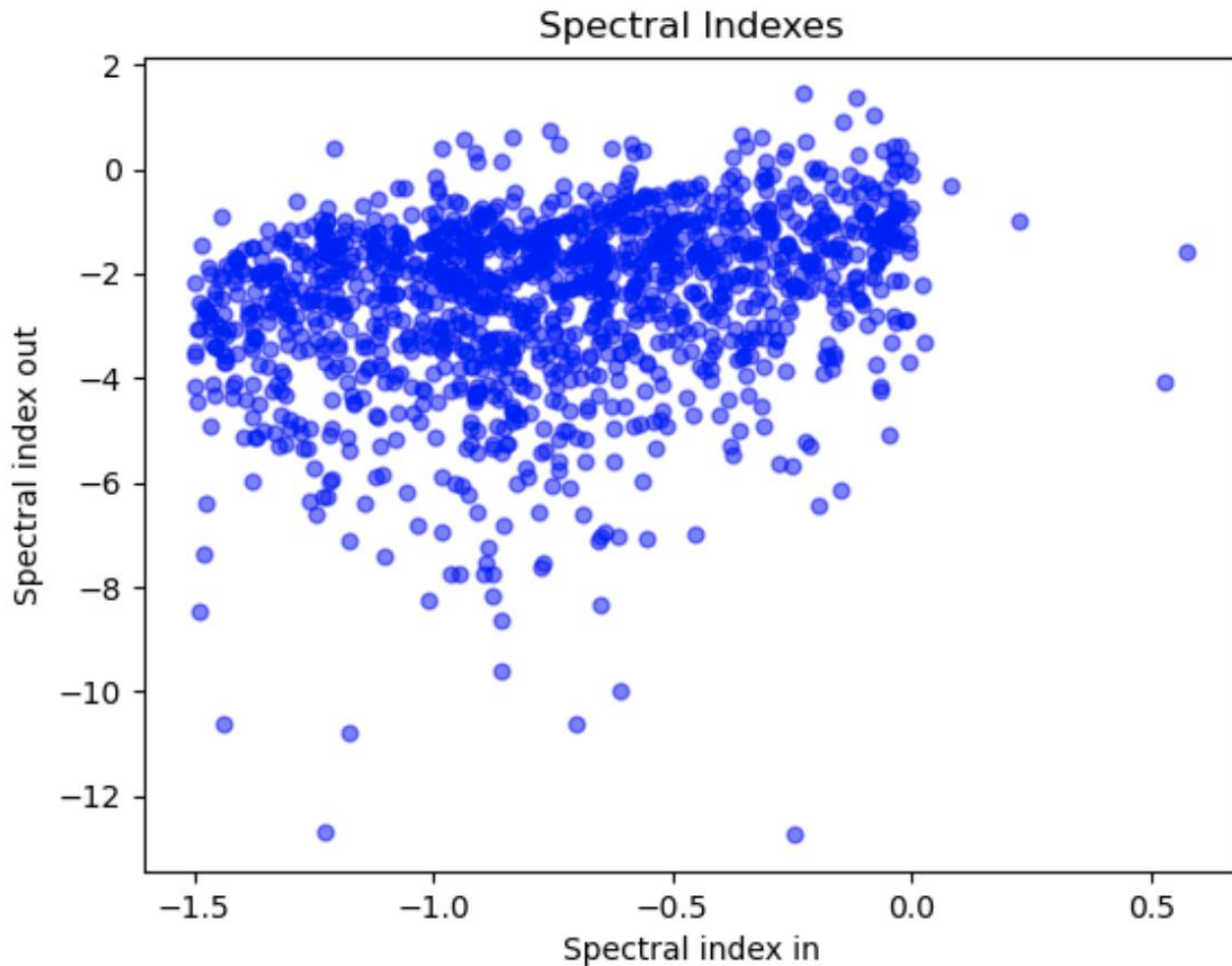
- 许多弱源没有被识别到
- 从右图可以看出越远离相位中心，源流量越小



主要是由于主波束校正对图像边缘的旁瓣不准确，导致源变弱了，从而弱源没被检测到



评价指标——光谱指数



测试星表和原始星表的源光谱指数比较

- 测试星表的源的光谱指数被过分估计了，即越远离参考频率的源流量会越不准确
- 进一步探究发现估计不太准确的源倾向位于图像边缘



- 主波束效应影响
- 成像算法对图像边缘的影响，导致源搜寻软件在图像边缘无法准确推断源



现有评价方法可改进之处

- 评价指标可以更加实用和更准确地评价成像质量
- 没有对源的偏振属性进行评价
- 源搜寻软件PYBDSF用于SKA是否足够准确
- 图像边缘质量不好可能是连续成像管线造成的，也可能是源搜寻软件在边缘表现不好，需要进行进一步探究
- 耗费时间较长，Celeste使用8个核心对32K*32K的图像进行质量评价，检测到并拟合20,000个源，共耗时35分钟



研究计划

- 如何提出更加准确的质量评价指标：
 - 深入探究连续成像管线可能会带来的影响，从而设计相关的指标来反映是否存在该影响、影响程度
 - 将偏振补充入质量评价体系
- 如何使得质量评价过程更加准确：
 - 验证PYBDSF对图像边缘估计背景是否准确
 - 验证其他源搜寻算法的表现——AEGEAN, BLOBCAT
- 如何使得质量评价过程更加快速
 - 抽样评价
 - 探究有没有更快速的算法



广州大学
Guangzhou University

感谢大家的聆听与
指导!