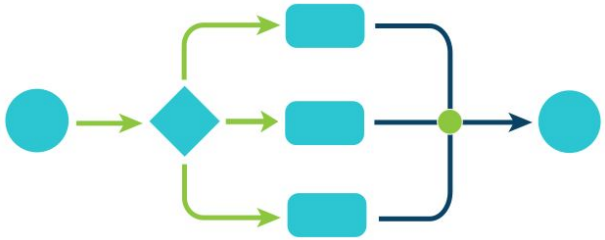
Scientific workflows

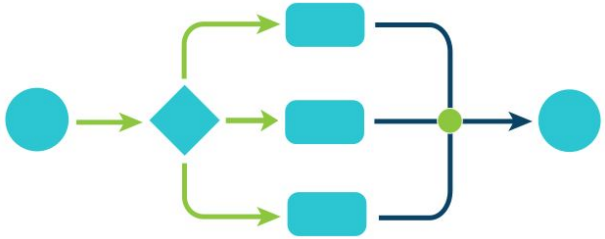


Adriano Pieres
Michel Aguena

Outline

- What is a workflow?
- Example I: ModStar
- Example II: WaZP

Workflow in Portuguese...



... uma **sequência de passos necessários para automatizar processos**, de acordo com um conjunto de regras definidas, permitindo que estes possam ser transmitidos de uma etapa para outra.

https://www.significados.com.br/workflow/

Example I: a closer look at ModStar workflow

Scientific motivation

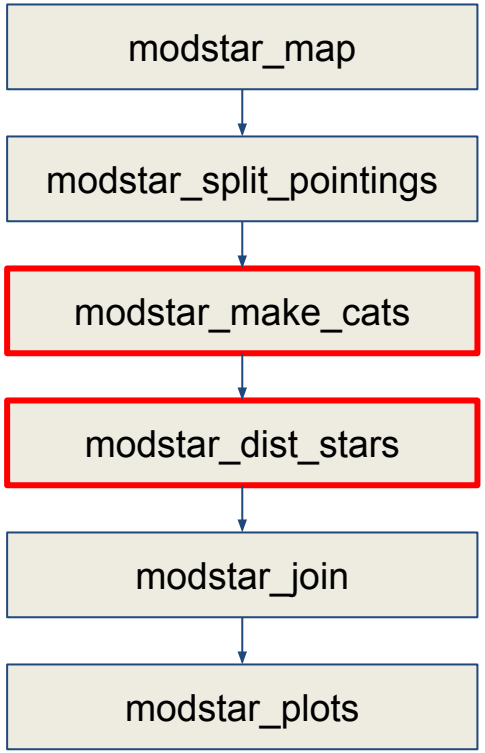
- Documentation: PD 002/2019
- des-portal.linea.gov.br > Pipelines > Science Analysis > Galaxy Archaeology > Modstar
- Similar to http://stev.oapd.inaf.it/cgi-bin/trilegal but with positions, a wider range of models and supporting large files

- Simulate the stellar content of a specific field with TRILEGAL models;
- Input: photometric system (DES and SDSS available), the coordinate system (Equatorial or Galactic) and the range in latitude and longitude in the specific coordinate system. The pipeline simulates contiguous or non contiguous fields (,):

- Non-contiguous region:
- LONMIN: 0, 100
- LONMAX: 5, 105
- LATMIN: 5, -80
- LATMAX: 10, -75

- Galactic model with parameters to each component;

ModStar workflow



Input: select data from user. Map regions and create files to run simulation (how many pixels, phot system, etc.)

Select how many and which pixels are going to which nodes

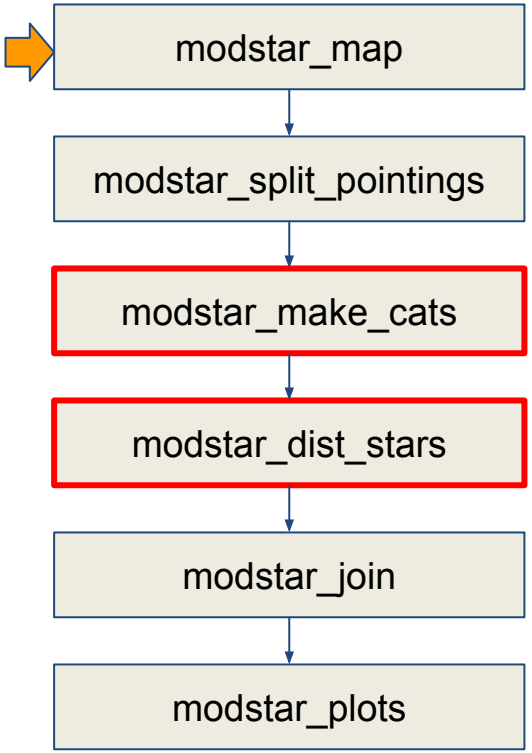
Run Trilegal to each pixel in each node.

Creates the position (RA and DEC / l and b) of each star.

Join all the healpixels in a single catalog (intended to download a single file)

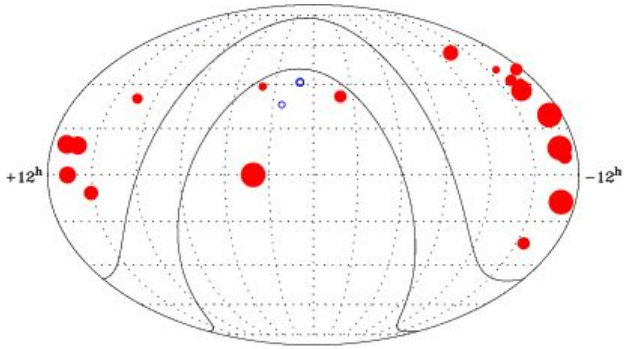
Run the plots

ModStar workflow



Which coordinates should be applied?

Equatorial: Crowded fields in low g alactic latitudes



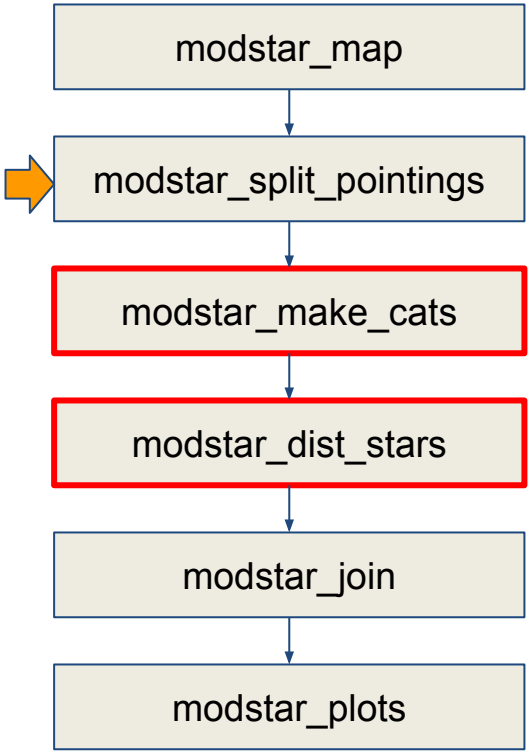
Solution: restrict area using vertices in LAT, LON

Select HPixels using query_polygon

Increase area due to geodesical lines connecting vertices

List all HealPixels to be simulated

ModStar workflow

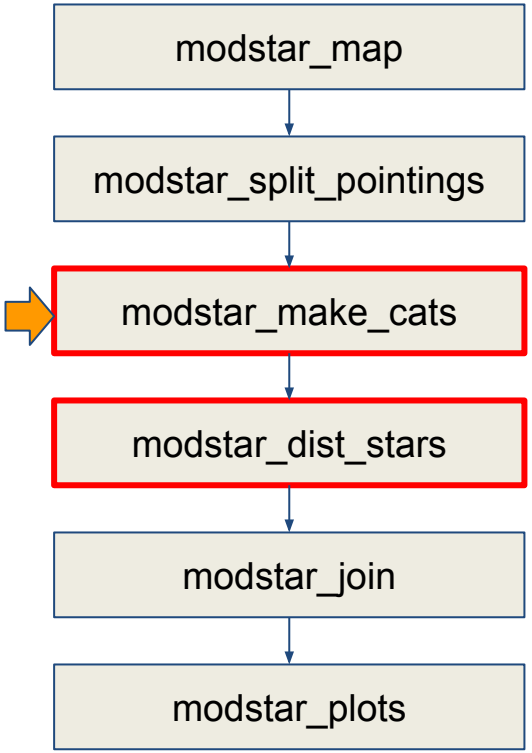


Divide HPixels in chunks

Error when nnodes > nfields

1 -> N

ModStar workflow

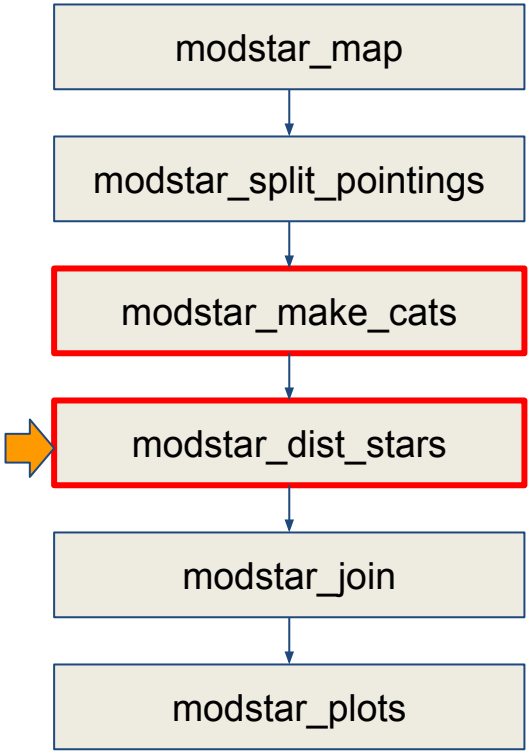


Run Trilegal code to each pointing

Trilegal simulates magnitudes in the specific photometric system, velocities (pmdec, pmracosdec, radial velocity), intrinsec features (mass, logg, effective temperature), distance.

Trilegal does not simulate position of the stars

ModStar workflow

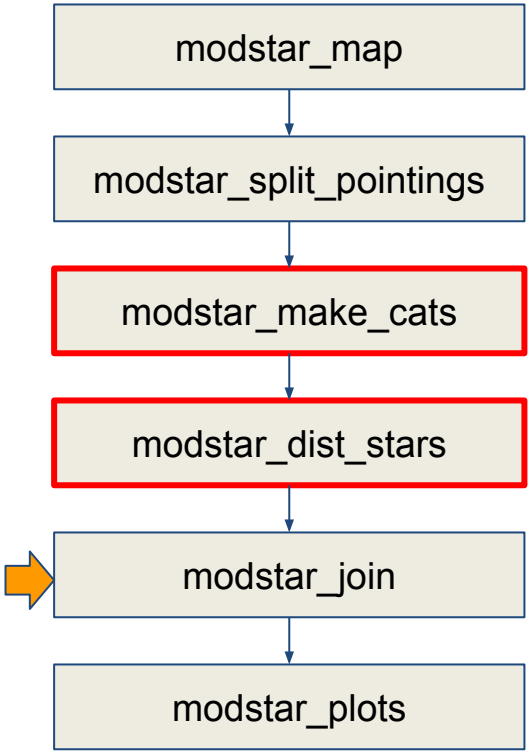


Dist_stars distribute stars in the specific HealPixel.

Pointings are simulated using HP nside = 64 = 1 square degree.

Using nested scheme and nside = 65536 (healpix side = 0.17 arcsec), the simulated positions are created.

ModStar workflow

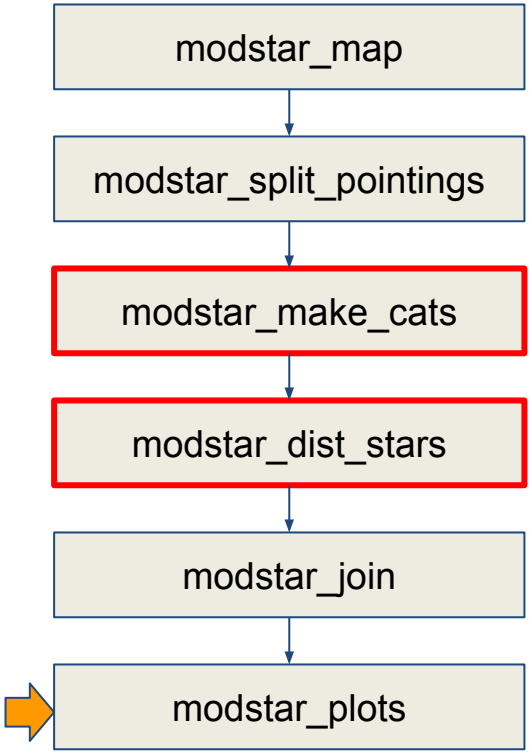


In the case the user selects the break in the power-law indice (Watkins et al. 2009; Deason et al. 2011; Sesar et al. 2011), that is done here.

Finally, the simulated positions are incorporated in the catalog from Trilegal in each HPixel

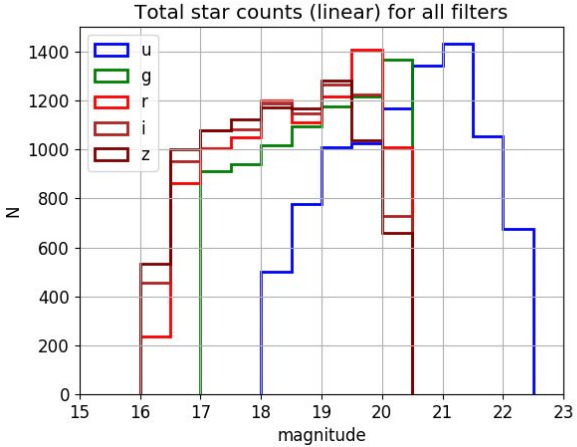
All the HPixels catalogs are joint in a single catalog intended to download a single file

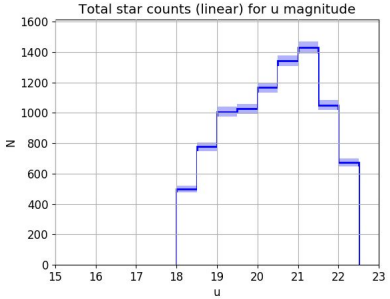
ModStar workflow



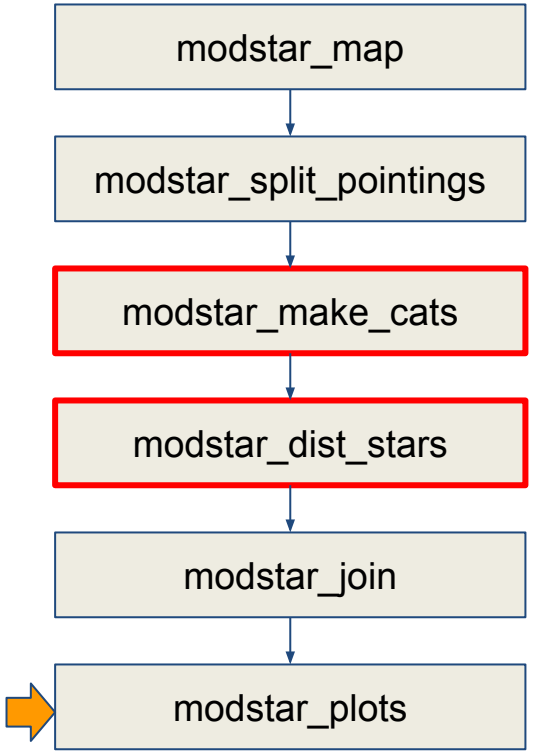
Plots showing the distribution of the stars in the simulated catalog are done here.

Example:
https://adriano.pieres.linea.gov.br/VP/getViewProcessCon?proces

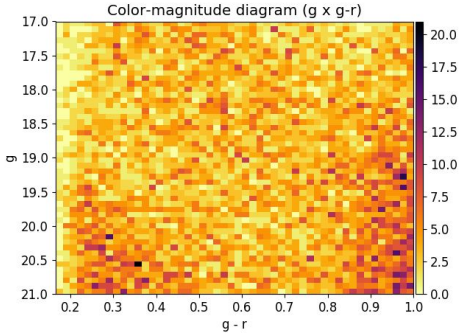


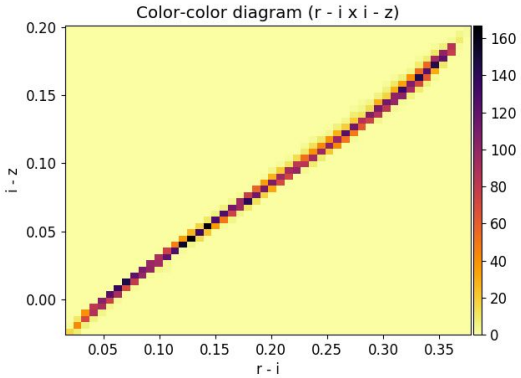


ModStar workflow

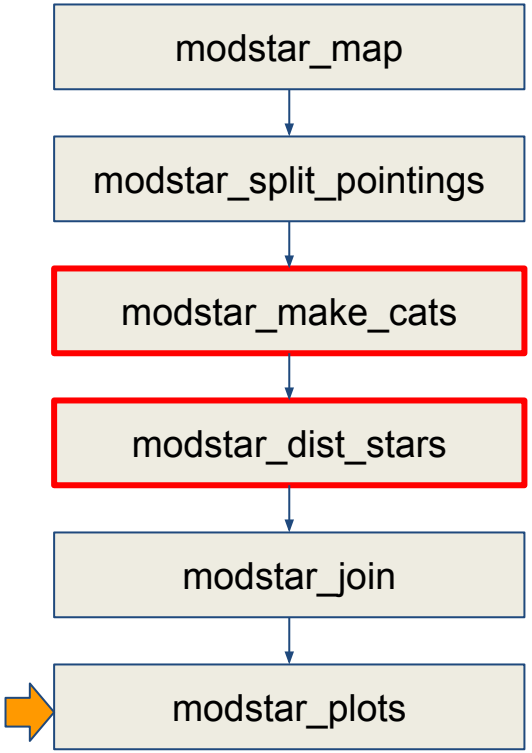


Plots showing the distribution of the stars in the simulated catalog are done here.

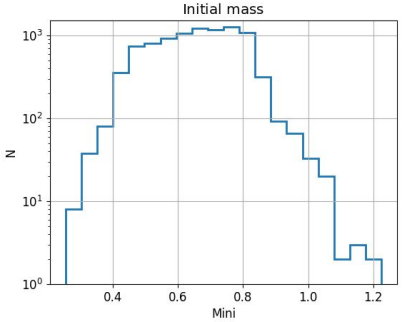


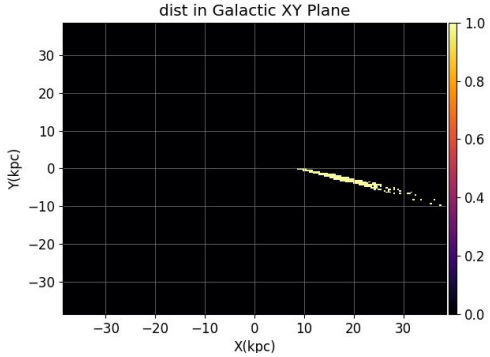


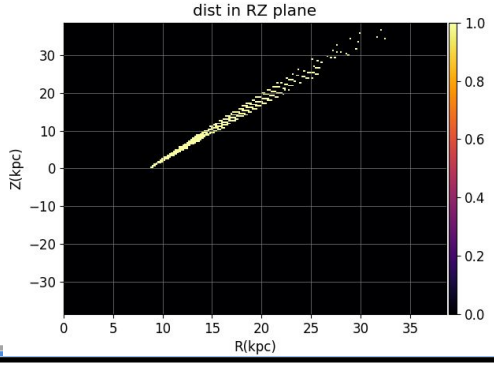
ModStar workflow

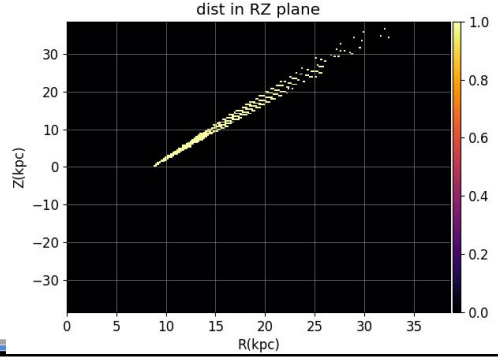


Plots showing the distribution of the stars in the simulated catalog are done here.

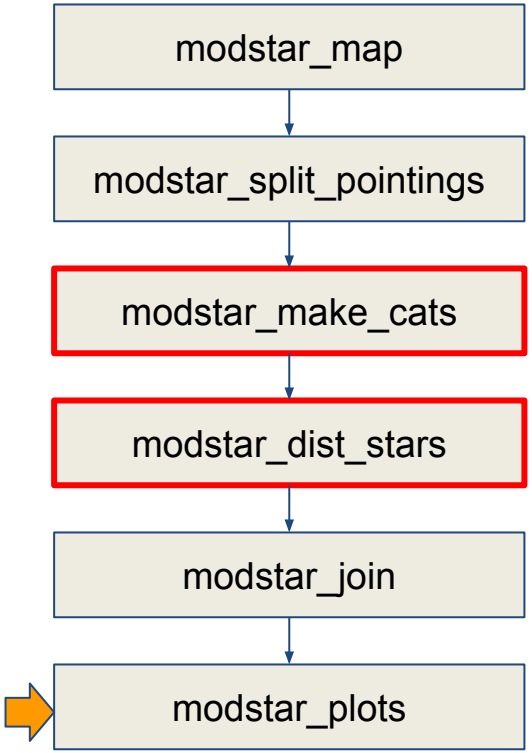




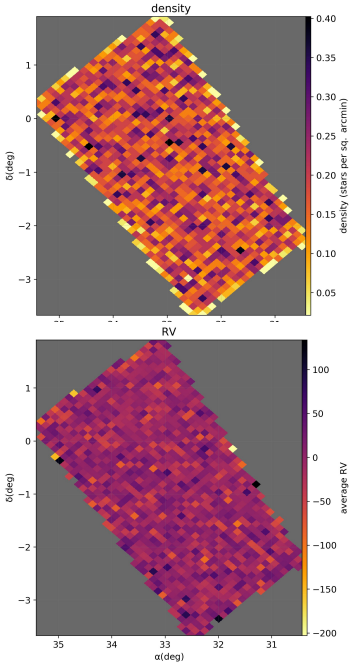


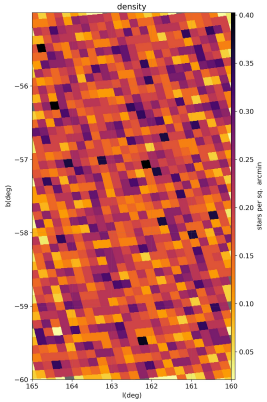


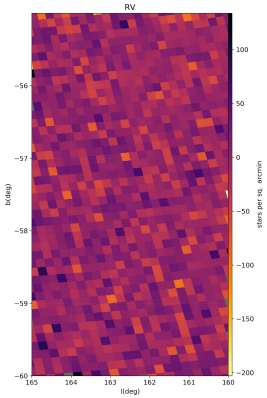
ModStar workflow



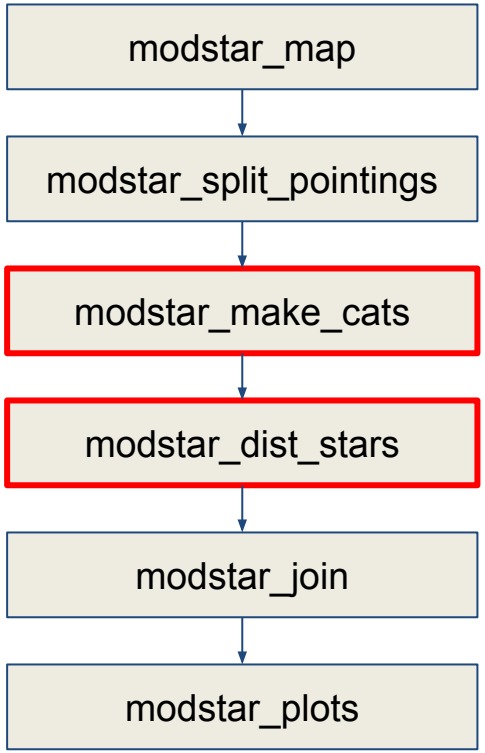
Plots showing the distribution of the stars in the simulated catalog are done here.







ModStar workflow - Download simulated catalog



- des-portal.linea.gov.br
- 'Dashboard'
- Select any Release and dataset
- Science workflows -> Click on the link available (usually a number) in the line starting with the Modstar
- Click on the links of products
- Click on 'Download'
- A link will sent to you to download the simulated catalog via ftp

ModStar workflow - Download simulated catalog

modstar_map

↓

modstar_split_pointings

↓

modstar_make_cats

↓

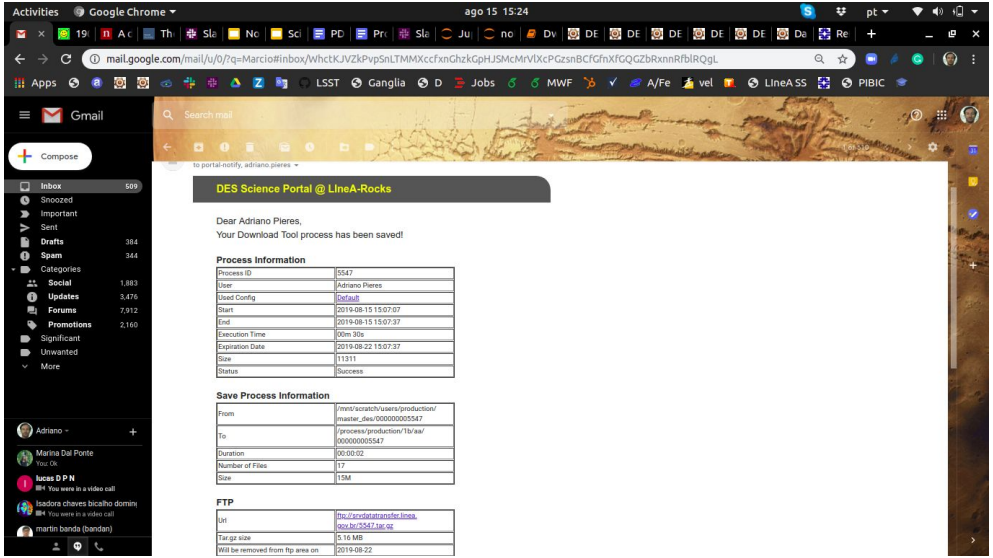
modstar_dist_stars

↓

modstar_join

↓

modstar_plots



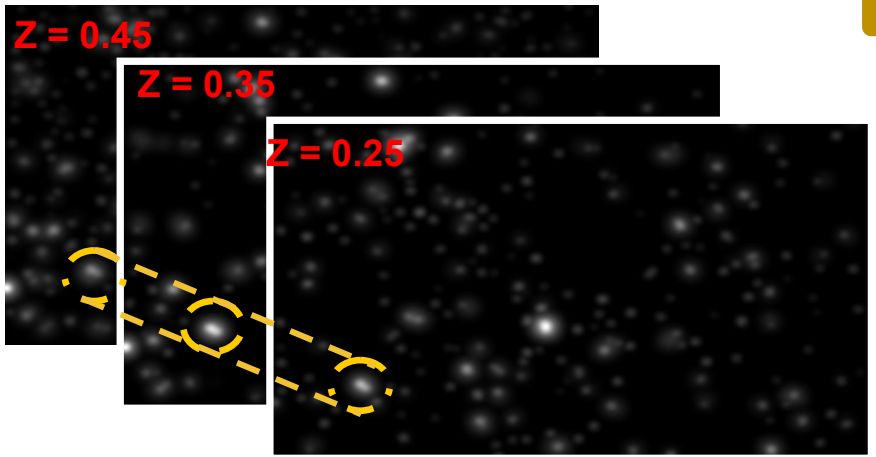
Example II: a closer look at WaZP

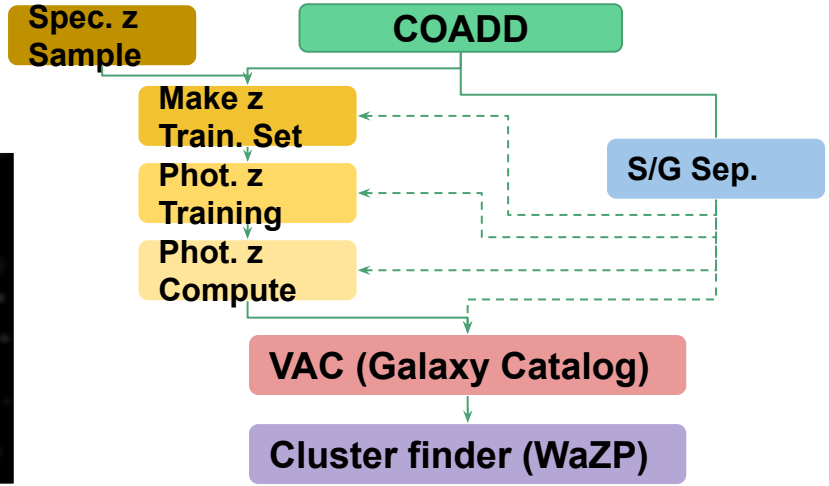
WaZP - Scientific motivation

- Detection of galaxy clusters based on galaxies with photometric redshifts
- Does not depend on red-sequence of galaxies

Location:

des-portal.linea.gov.br > Pipelines > Science Analysis > Cluster > WAZP

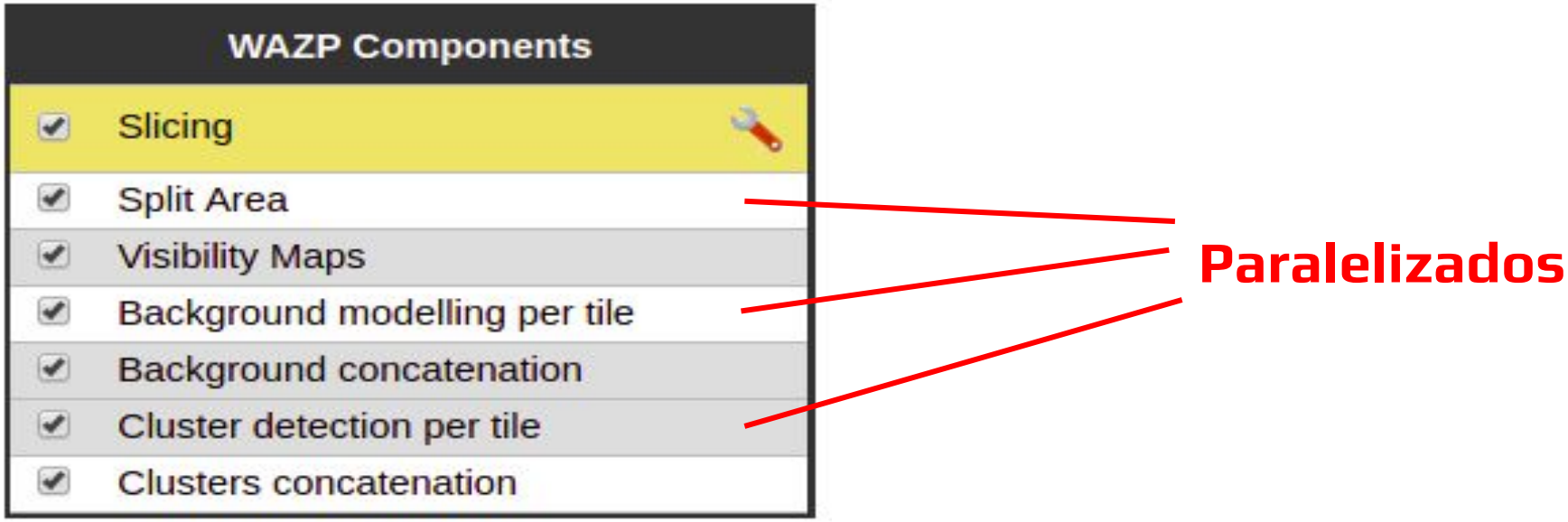




WaZP workflow

WAZP Components	
☑ Slicing	→ portal/des/components/wazp_slice_zmag
☑ Split Area	→ portal/des/components/wazp_split_area
☑ Visibility Maps	→ portal/des/components/wazp_visibility_maps
☑ Background modelling per tile	→ portal/des/components/wazp_bkg_model
☑ Background concatenation	→ portal/des/components/wazp_pre_tile
☑ Cluster detection per tile	→ portal/des/components/wazp_tile
☑ Clusters concatenation	→ portal/des/components/wazp_concatenate

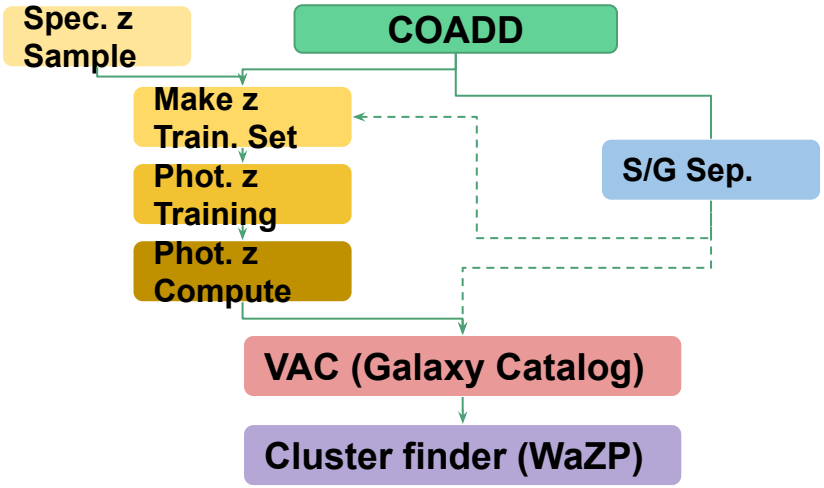
WaZP workflow



WaZP inputs

Provenance

Name	Process ID
WAZP	5239
Cluster	5211
Install Catalogs	139
Photo-z Compute	5100
Install Catalogs	139
Photo-z Training	5042
Install Catalogs	469
Install Catalogs	35
Install Catalogs	472
Install Catalogs	139
Install Catalogs	119
Training Set Maker	3529
SG Separation	2584
Install Catalogs	469
SG Separation	2437
Install Catalogs	35
Install Catalogs	472
SG Separation	3525
SG Separation	2619
Install Catalogs	139
SG Separation	3312
Install Catalogs	119
Spectroscopic Sample	2953
SG Separation	2813



WaZP run

Process Summary | Results | Comments

INPUT

Input Data	
Release	Y1A1
Dataset	SDT
Value-Added Catalogs	Cluster 44

Additional Information	
Zp correction	SFD98
Extinction map	SFD98
Footprint area:	1380.30 sq deg
Mean density:	8.50 objects/(sq arcmin)
Pipeline name:	VAC Cluster
Photo-z	DNF
S/g classification	Y1 Modest v2

Process Information	
Stage	Science Analysis
Process ID	5239
User	Christophe Benoist
Used Config	Custom
Start	2019-06-11 04:59:19
End	2019-06-11 07:08:54
Execution Time	2h 09m 35s
Expiration Date	2019-06-18 07:08:54
Size	260645492
Status	Success
Code Viewer	
Overall Success Rate	100%
Total Number of Jobs	752
Time Profiler	

Output Data	
Targets	Cluster Members 56
Targets	Galaxy Clusters 56
Targets	Cluster Members 55
Targets	Galaxy Clusters 55

Module	Duration	Config	Error Log	Pipeline Out	Log	Condor Log	NC	Success Rate	Status
Slicing	0:00:28						null	-	✔
Split Area	0:06:37	-					null	-	✔
Visibility Maps	0:15:58	-	-	-	-	-		100%	✔
Background modelling per tile	0:19:37	-	-	-	-	-		100%	✔
Background concatenation	0:02:35	-					null	-	✔
WAZP split per tile	0:00:05	-					null	-	✔
Cluster detection per tile	0:22:13	-	-	-	-	-		100%	✔
Clusters concatenation	0:43:12	-					null	-	✔

WaZP run

Process Summary | Results | Comments

Input Data	
Release	Y1A1
Dataset	SPT
Value-Added Catalogs	Cluster 44

Additional Information	
Zp correction	SFD98
Extinction map	SFD98
Footprint area:	1380.30 sq deg
Mean density:	8.50 objects/(sq arcmin)
Pipeline name:	VAC Cluster
Photo-z	DNF
S/g classification	Y1 Modest v2

Process Information	
Stage	Science Analysis
Process ID	5239
User	Christophe Benoist
Used Config	Custom
Start	2019-06-11 04:59:19
End	2019-06-11 07:08:54
Execution Time	2h 09m 35s
Expiration Date	2019-06-18 07:08:54
Size	260645492
Status	Success
Code Viewer	
Overall Success Rate	100%
Total Number of Jobs	752
Time Profiler	

OUTPUT

Output Data	
Targets	Cluster Members 56
Targets	Galaxy Clusters 56
Targets	Cluster Members 55
Targets	Galaxy Clusters 55

Module	Duration	Config	Error Log	Pipeline Out	Log	Condor Log	NC	Success Rate	Status
Slicing	0:00:28						null	-	
Split Area	0:06:37	-					null	-	
Visibility Maps	0:15:58	-	-	-	-	-		100%	
Background modelling per tile	0:19:37	-	-	-	-	-		100%	
Background concatenation	0:02:35	-					null	-	
WAZP split per tile	0:00:05	-					null	-	
Cluster detection per tile	0:22:13	-	-	-	-	-		100%	
Clusters concatenation	0:43:12	-					null	-	

WaZP run

Process Summary | Results | Comments

Input Data	
Release	Y1A1
Dataset	SPT
Value-Added Catalogs	Cluster 44

Additional Information	
Zp correction	SFD98
Extinction map	SFD98
Footprint area:	1380.30 sq deg
Mean density:	8.50 objects/(sq arcmin)
Pipeline name:	VAC Cluster
Photo-z	DNF
S/g classification	Y1 Modest v2

Process Information	
Stage	Science Analysis
Process ID	5239
User	Christophe Benoist
Used Config	Custom
Start	2019-06-11 04:59:19
End	2019-06-11 07:08:54
Execution Time	2h 09m 35s
Expiration Date	2019-06-18 07:08:54
Size	266045492
Status	Success
Code	
Overall Success Rate	100%
Total Number of Jobs	752
Time Profiler	

DURATION

Output Data	
Targets	Cluster Members 56
Targets	Galaxy Clusters 56
Targets	Cluster Members 55
Targets	Galaxy Clusters 55

Module	Duration	Config	Error Log	Pipeline Out	Log	Condor Log	NC	Success Rate	Status
Slicing	0:00:28						null	-	✓
Split Area	0:06:37	-					null		✓
Visibility Maps	0:15:58	-	-	-	-	-		100%	✓
Background modelling per tile	0:19:37	-	-	-	-	-		100%	✓
Background concatenation	0:02:35	-					null		✓
WAZP split per tile	0:00:05	-					null	-	✓
Cluster detection per tile	0:22:13	-	-	-	-	-		100%	✓
Clusters concatenation	0:43:12	-					null		✓

PARALLEL

WaZP run

Cluster WAZP

Process ID: 5239

Process Summary | **Results** | Comments

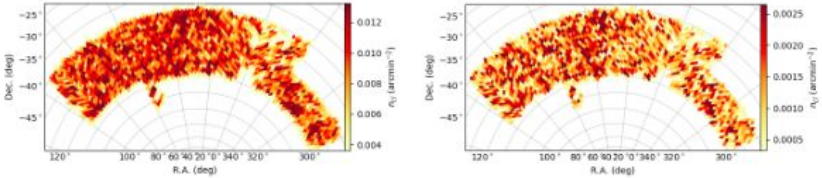
Summary | WAZP Slices | WAZP BKG Model | **WAZP Results**

Clusters | Members

Quick Navigation

- Satial distribution
- Properties Distribution
- NGALS
- Background
- Other

Satial distribution ?



WaZP run

Cluster WAZP

Process ID: 5239

Process Summary | **Results** | Comments

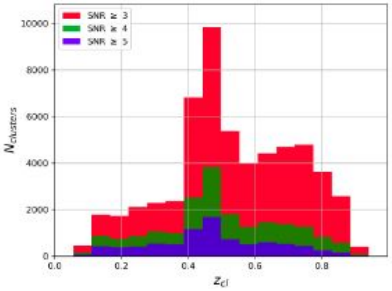
Summary | WAZP Slices | WAZP BKG Model | **WAZP Results**

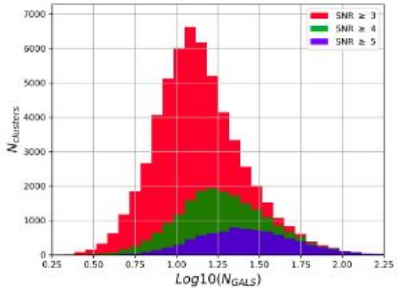
Clusters | Members

Quick Navigation

- Satial distribution
- Properties Distribution
- NGALS
- Background
- Other

Properties Distribution ?





WaZP run

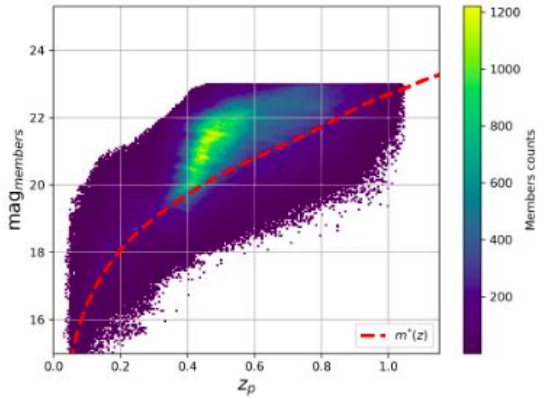
Cluster WAZP

Process ID: 5239

Process Summary | **Results** | Comments

Summary | WAZP Slices | WAZP BKG Model | **WAZP Results**

Clusters | **Members**



Questões

Obrigado