



The Industrial Revolution of Galaxies Cluster Finding in the Dark Energy Survey

Ricardo Ogando
ogando@linea.gov.br

OROTOUR GARDEN HOTEL • RUA ENGº GUSTAVO KAISER, 165 • VILA NATAL - CAMPOS DO JORDÃO, SP • CEP 12460-000

SCIENCE WITH THE LSST: A BRAZILIAN/US JOINT WORKSHOP

APRIL 1-4, 2012



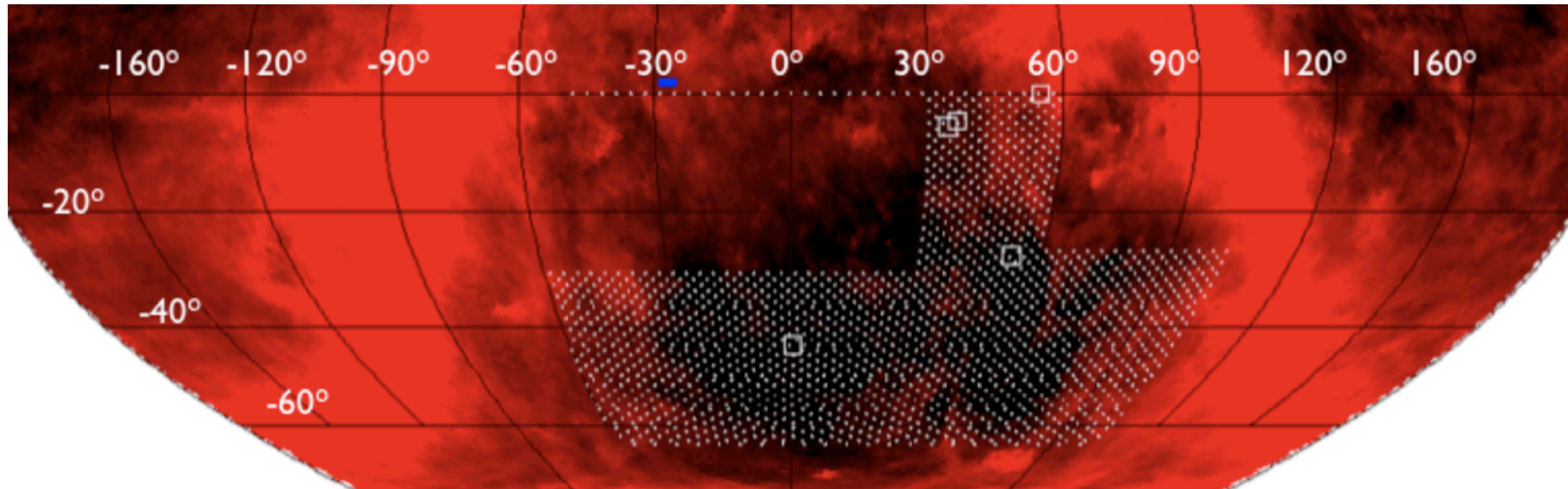


Outline



- Introduction
- Cluster Finders Industrialization
- Data handling
- Parallelization
- Results
- Scientific collaboration network
- Fine tuning and Cluster finding challenge
- Lessons learned

- Dark Energy Survey 5000 square degrees
- 400 million galaxies spread on the cosmic web
- Four probes to cosmology: SN, BAO, WL and **Clusters**
- Several methods to find clusters of galaxies
- Understand their selection functions → better cosmology.





Cluster Finders Industrialization



- Run Cluster Finders (CFs) on **mock** catalogs
- Match **clusters** to **halos** in simulations
 - Calculate **completeness** and **purity**
- Have several CFs in a controlled environment
 - Keep track of data provenance, code version, and configurations
- Faster processing response time on data releases



Cluster Finders Industrialization



- **Matched Filter** (L. Olsen), **WAZP** (C. Benoist), **VTRS** (W. Barkhouse), **GMBCG** (J. Hao), **Lambda** (E. Rykoff), **C4+** (L. Baruah), and **zVT** (M. Soares-Santos)
- **Framework** to compare results allowing:
 - **Fine tuning.** Finding the best set of CF input parameters, e.g. M^* and threshold.
 - **Cluster Finding challenge.** Pros and cons of each CF method. Assess their selection functions.
- **Compliance:** CFs produce common output: RA, Dec, z, mass proxy rank



The Cluster Finder pipeline



The Dark Energy Survey

Ricardo Ogando
Welcome to the Ogando portal.

Home Release Notes My Workspace Pipelines Tools Data Server Documentation Help Credits Other Portals Logout

Introduction to the Science Portal

This portal is a web-based service that gives the user access to the portal, monitor the progress of the survey, and to manage products. Links and documentation are available in the portal with in-line help available in various parts of the system as well as short explanatory demonstration movies available in the portal also provides tools for power-users and system administrators to oversee the entire system.

The services currently available under the menu folder include:

- **My Workspace:** provides access to the profile, processes, configuration status, allowing the access to: 1) the portal machines through ssh; 2) the portal also provides tools for power-users and system administrators to oversee the entire system.
- **Pipelines:**
 - **Data Reduction:** provides access to Quick Reduce, a tool to reduce data and will turn into the PRECam reduction pipeline.
 - **Advanced Products:** here the user can produce and export catalogs, either regular or value-added, and validate them using the tools available in the portal.
 - **Science:** provides access to the available scientific pipelines, grouped in sub-categories (sub-folders) representing different working and study groups. The available pipelines (processes) in each group varies in the degree of sophistication and scope but all consist of concatenated full configurable modules chosen according to the user-specified workflow and algorithms, that query for data, process one or more jobs, depending on the number of input data sets, and generate scientific products that are stored in a temporary directory, which can later be saved or exported by the user. The processes can be run offline with the users being notified by e-mail upon completion.
- **Tools:** provides a large variety of tools, like processes running in the cluster, code manager, cluster environmental information, enabling the access to the administrative layer of the portal for those with permission. Here is also available the means to monitor the Survey Coverage
- **Data Server:** provides the infrastructure and the tools to upload data to the portal and to export the data from within the portal framework to a user specified target

Science Portal

Cluster Finder

Cluster Comparison

Cluster

SN

WL

Photo-z

Simulation

Galaxy Archeology

Galaxy Evolution

QSO

Strong Lensing

Combined Probes

LSS



Code versioning and share



My Workspace

My Processes | My Comments | My Space | My Uploads | My Export | My Configurations | **My Developer Zone** | My Profile

Filter: Codes

Show 10 entries

First | Previous | 1 | Next | Last

Code	Creation Date	Description	Edit Permission	Remove Code
des_cfc	2010-07-15 17:57:05	Cluster Finder Comp...		

New Code

Add Permission

Groups | Users

User with permission

Angelo Fausti	<input checked="" type="radio"/> Read	<input type="radio"/> Write	
Eli Rykoff	<input type="radio"/> Read	<input checked="" type="radio"/> Write	
Jeeseon Song	<input checked="" type="radio"/> Read	<input type="radio"/> Write	
Jiangang Hao	<input checked="" type="radio"/> Read	<input type="radio"/> Write	
Leon Baruah	<input checked="" type="radio"/> Read	<input type="radio"/> Write	
Marcelle Soares-Santos	<input checked="" type="radio"/> Read	<input type="radio"/> Write	
Ricardo Ogando	<input type="radio"/> Read	<input checked="" type="radio"/> Write	

Add Permission in user: Adam Hawken

Apply Permission

Code is versioned in My Developer Zone and shared among collaborators with different permissions

- CF codes wrapped in **python** modules
- Modules concatenated in **XML** workflows

Workflow

Retrieving Data

☒ Data Organizer

☒ Data Retriever

Algorithm

☒ GMBCG

☒ Consolidator

GMBCG
 Matched Filter
 WAZP
 VT RS

Analysis

☒ Global Generic plots

☒ Membership Matching

☒ QA Plots

☒ Science Requirements

☐ Save process
☐ Transfer process

Save Config

Workflow

Retrieving Data

☒ Data Organizer

☒ Data Retriever

Algorithm

☒ Structure Detection

☒ Masking

☒ Filter Overlaps

☒ Structure Analysis

☒ Membership

☒ Merge Results

Analysis

☒ Global Generic plots

☒ Membership Matching

☒ QA Plots

☒ Science Requirements

The Dark Energy Survey

Ricardo Ogando
Welcome to the Ogando portal.

Home Cluster Finder Data Configuration Documentation Help

Provenance Search

Refine by **Hint** **Option** **Add/Remove**

Coordinates Plot Footprint

RA: 20 deg
DEC: 40 deg
Boxsize: 2 deg

Execute Query

3 Box search

Provenance Search

Refine by **Option** **Add/Remove**

Coordinates

Execute Query

2 Choose constraint

Provenance Search

Location **Input Data** **Type**

ON Addgals v3.02 Catalog

Addgals v3.02
Addgals v3.04
DC6B
DC6B3
Uploads

1 Select data



Data handling



- ADDGALS: mock catalogs (200sqdeg)
- Data Challenges: mock observations (100sqdeg)
- Blind Cosmology Challenge: unknown cosmology mock catalog (5000sqdeg)
- Other data: CFHT, S82, and BOSS
- Ingested to PostgreSQL database.
- Q3C indexing for large astronomical catalogs

Provenance Search Selection					
Found just one page.					
File Name	Upload Date	File Type	File Class	Uploaded by	
<input type="checkbox"/> redmap_v2.13	2011-06-28 12:45:04.847170	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> lambda220	2011-06-17 22:49:18.354475	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> testcat	2011-06-06 14:07:04.167049	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> DES_truth_2n3n4.fit	2011-06-05 17:18:04.953233	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> DES_truth_2n3n4.fit	2011-06-05 17:18:04.952156	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> DES_truth_2n3n4.fit	2011-06-05 17:18:04.954176	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> Lindsey_total	2011-04-23 22:29:51.838825	application/fits-table	science	Jeeseon Song	
<input type="checkbox"/> Lindsey_new	2011-04-21 00:34:59.220673	application/fits-table	science	Jeeseon Song	



Provenance Search Selection		
File Name		
<input type="checkbox"/> C4 v2.11 final catalog		
<input type="checkbox"/> C4 v2.11 test catalog		

CFOs can also upload their cluster catalogs

Uploaded by	
Leon Baruah	
Leon Baruah	



Provenance Search Selection					
Found just one page.					
File Name	Upload Date	File Type	File Class	Uploaded by	
<input type="checkbox"/> rykoff_v2.13_matched_bgerke	2011-06-02 12:38:02.801418	application/fits-table	science	Brian Gerke	
<input type="checkbox"/> hao_gauss_v211_testing	2011-03-16 17:01:01.627407	application/fits-table	science	Brian Gerke	
<input type="checkbox"/> rykoff_v2.13_clusters	2011-02-17 20:32:08.078426	application/fits-table	science	Brian Gerke	
<input type="checkbox"/> baruah_v2.11_matched	2011-02-17 17:34:44.430108	application/fits-table	science	Brian Gerke	
<input type="checkbox"/> rykoff_v2.11_matched	2011-02-16 00:44:01.064198	application/fits-table	science	Brian Gerke	

Provenance Search Selection					
Found just one page.					
File Name	Upload Date	File Type	File Class	Uploaded by	
<input type="checkbox"/> S82_redMaPPer_v3.4_nord2	2011-09-10 19:02:37.614415	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> S82_redMaPPer_v3.4_nord	2011-09-10 17:20:22.655624	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> S82_redMaPPer_v3.4_like	2011-09-10 17:03:26.664036	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> Modprop Halos	2011-09-09 12:32:44.750945	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> Constant Halos	2011-09-09 12:30:54.903871	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> Bowl Halos	2011-09-09 12:30:11.745120	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> Bowl-gr Halos	2011-09-09 12:29:50.924805	application/fits-table	science	Eli Rykoff	
<input type="checkbox"/> Baseline Halos	2011-09-08 13:29:26.451147	application/fits-table	science	Eli Rykoff	



Select All Unselect All Select All Pages Unselect All Pages There are 8 items returned.

Confirm

- Important for large area analysis
- (almost) Embarrassingly parallel
- Resolve overlaps by bisection before stitching
- Some CF needs post-processing

Data Organizer

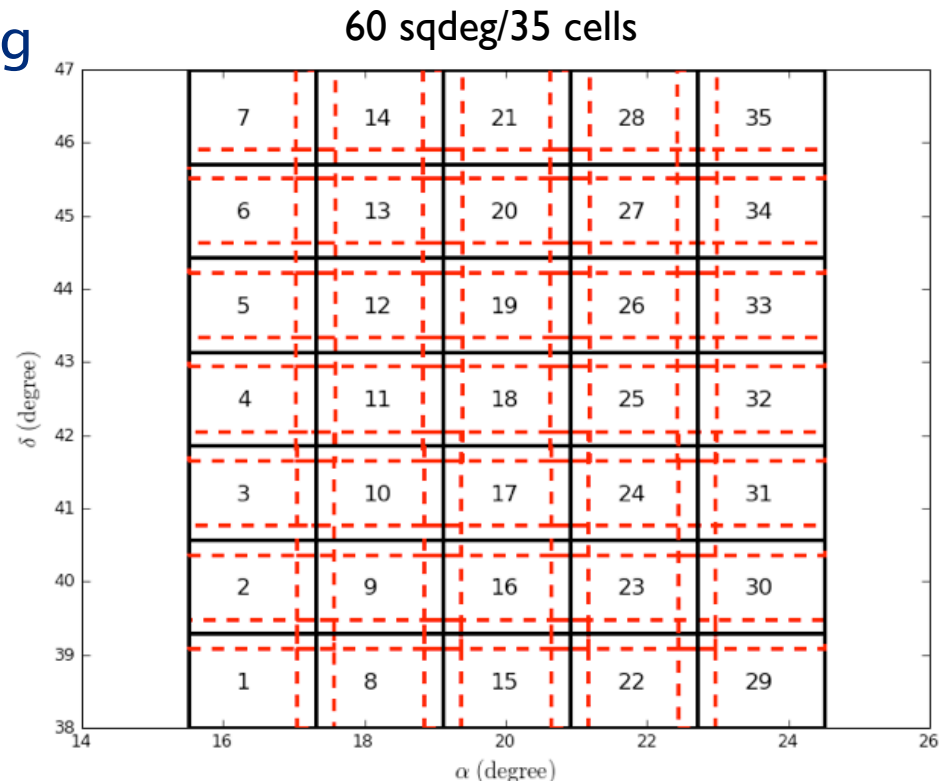
Parallelization

Parameters

Area of individual cell (square degree)

Overlap between cells (degree)

Remove slim cells that do not fit area? ☒ Yes ☐ No



cell and overlap size configurable

Cluster Finder (WAZP)

View Processes

Results

Comments

Pipeline: Cluster Finder

User	Ricardo Ogando
Process	10007936
Start	2012-02-16 02:17:10
End	2012-02-16 08:50:15
Duration	06:33:05
Ingest Time	00:00:02
Expiration	2012-02-21 08:50:15
Jobs	78
Automatically Saved	Yes
Algorithm	WAZP
Configuration	View
Code Version	View

Input Data

Input data	Addgals v3.04
Type	Catalog
RA (deg)	20
DEC (deg)	42.5
Selected Region (deg)	9
Objects	4803125
Area (sq deg)	54.67
Density (objects/sq arc min)	24.4

Module	Duration	Products	Config	Error Log	Pipeline Out	Log	Condor Log	NC	Status
dataorganizer	00:02:05	---						nc01	✓
dataretriever	00:01:35	---	---	---	---	---	---		✓
Structure Detection	00:01:24	---		---	---	---	---		✓
Masking	00:00:01	---		---	---	---	---		✓
wazp_filter_overlaps	00:00:00	---	---	---	---	---	---		✓
Structure Analysis	02:26:21	---		---	---	---	---		✓
wazp_membership	00:32:14	---	---	---	---	---	---		✓
cl_consolidator	00:03:31		---					nc01	✓
Global Standard Plots	00:00:12	---	---					nc01	✓
Cluster Matching	00:05:36							nc01	✓
cl_cmp_qa_plots	00:03:29	---						nc01	✓
cluster_sci_req	00:00:08	---						nc01	✓

View Processes

Results

Comments

Structure Detection



Show 10 entries

Total entries: 12

First Previous 1 2 Next Last

#	Duration	Error Log	Pipeline Out	Log	Condor Log	NC Ip	Product	Status
1	00:00:55					nc14		✓
2	00:01:19					nc24		✓
3	00:01:31					nc05		✓
4	00:01:18					nc23		✓
5	00:01:15					nc10		✓
6	00:01:43					nc02		✓
7	00:01:10					nc16		✓
8	00:01:07					nc12		✓
9	00:01:56					nc26		✓
10	00:02:32					nc25		✓

Ok

dataorg									✓
dataret									✓
Structu									✓
Masking	00:00:01	---		---	---	---	---		✓
wazp_filter_overlaps	00:00:00	---	---	---	---	---	---		✓
Structure Analysis	02:26:21	---		---	---	---	---		✓
wazp_membership	00:32:14	---	---	---	---	---	---		✓
cl_consolidator	00:03:31		---					nc01	✓
Global Standard Plots	00:00:12	---	---					nc01	✓
Cluster Matching	00:05:36							nc01	✓
cl_cmp_qa_plots	00:03:29	---						nc01	✓
cluster_scl_req	00:00:08	---						nc01	✓

Cluster Finder (WAZP)

View Processes Results Comments

Summary Data Organizer Consolidator Generic Plots Cluster Comparison QA View unmatched CWG tests Science Requirements Matching metrics

Quick Navigation

[Input data](#)
[Data Organizer](#)
[Algorithm](#)
[Clusters properties](#)
[Halos properties](#)
[Halos properties within clusters](#)
[redshift limits](#)
[Generic plots](#)
[All halos properties](#)
[Cluster comparison: Truth Table](#)
[Cluster comparison](#)
[Cluster comparison \(statistics\)](#)
[Science Requirements](#)
[Matching metrics](#)

Input data

Data	Addgals v3.04
Type	Catalog
RA center (degree)	20
Dec center (degree)	42.5
Boxsize	9
Number of objects	4803125
Number of cells	12
Area of individual cell (square degree)	6.00
Overlap between cells (degree)	0.20
RA min (degree)	15.500
RA max (degree)	24.500
Dec min (degree)	38.000
Dec max (degree)	47.000
Total area (square degree)	54.67
Redshift type	Gaussian error photometric redshift
Magnitude type	ADDGALS OBSERVED magnitude

Data Organizer

Total area of the field (sq.deg.)	59.63
Mean area of tiles including overlap	6.00
Mean area of tiles without overlap	4.20
Ideal number of tiles	14.20
Effective Nr. of tiles	12
Min/Max areas of tiles accepted range	0.25 / 30.00
Min/Max areas of tiles including overlaps	5.62 / 7.02
Min/Max areas of tiles without overlaps	4.70 / 5.24

Algorithm

Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)

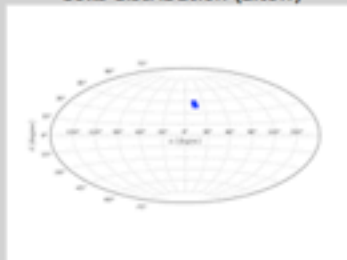
Quick Navigation

[Maps](#)[Footprint](#)[Irregular tiling](#)

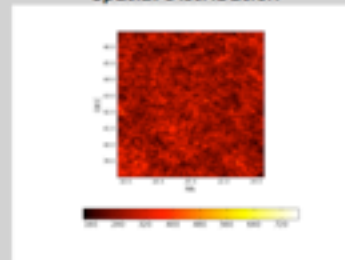
Maps

?

Cells distribution (altoff)



Spatial Distribution



Footprint

?

Cells distribution

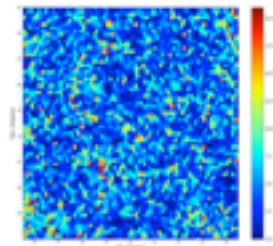


Cluster Finder (WAZP)

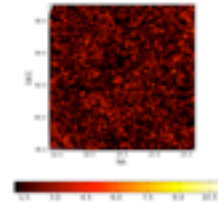
[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)**Consolidated catalog**

?

Consolidated Clusters heatmap



Consolidated Clusters heatmap



Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)[Clusters/Halos properties](#)[Clusters properties in redshift shells](#)

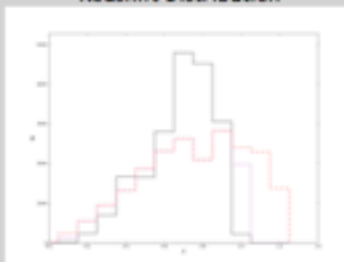
Quick Navigation

[Clusters properties](#)[Halos properties](#)

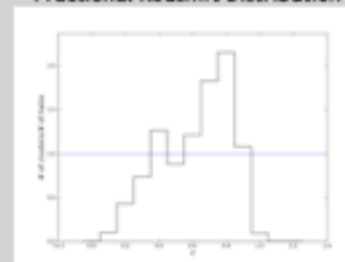
Clusters/Halos properties ?

Clusters properties ?

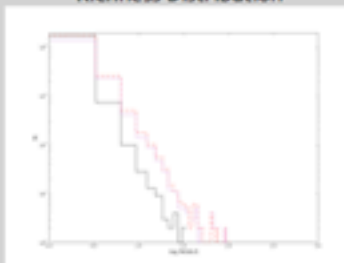
Redshift Distribution



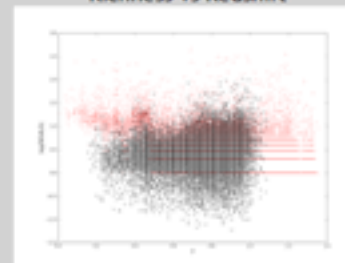
Fractional Redshift Distribution



Richness Distribution



Richness vs Redshift



Halos properties ?

Mass Distribution



M200 vs NGALS

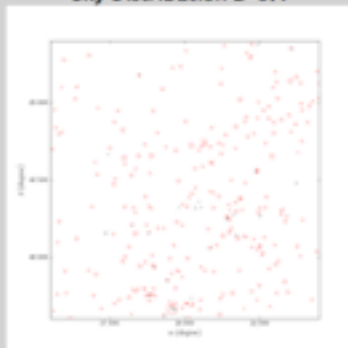


Cluster Finder (WAZP)

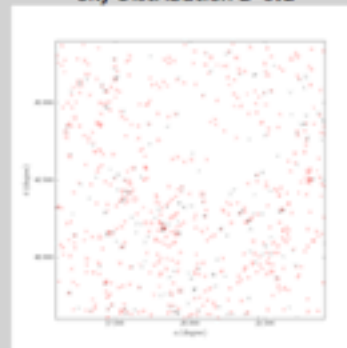
[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)[Clusters/Halos properties](#)[Clusters properties in redshift shells](#)

Spatial distribution

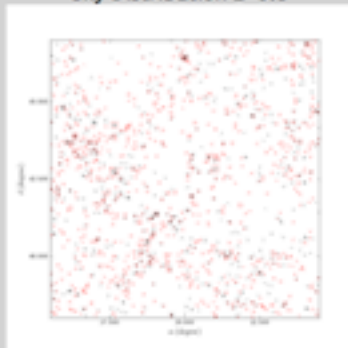
Sky Distribution $z=0.1$



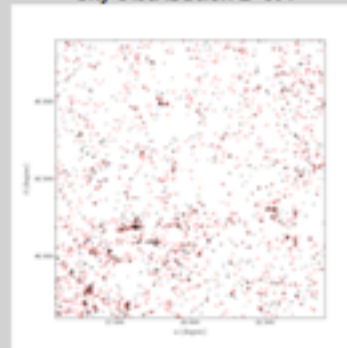
Sky Distribution $z=0.2$



Sky Distribution $z=0.3$



Sky Distribution $z=0.4$



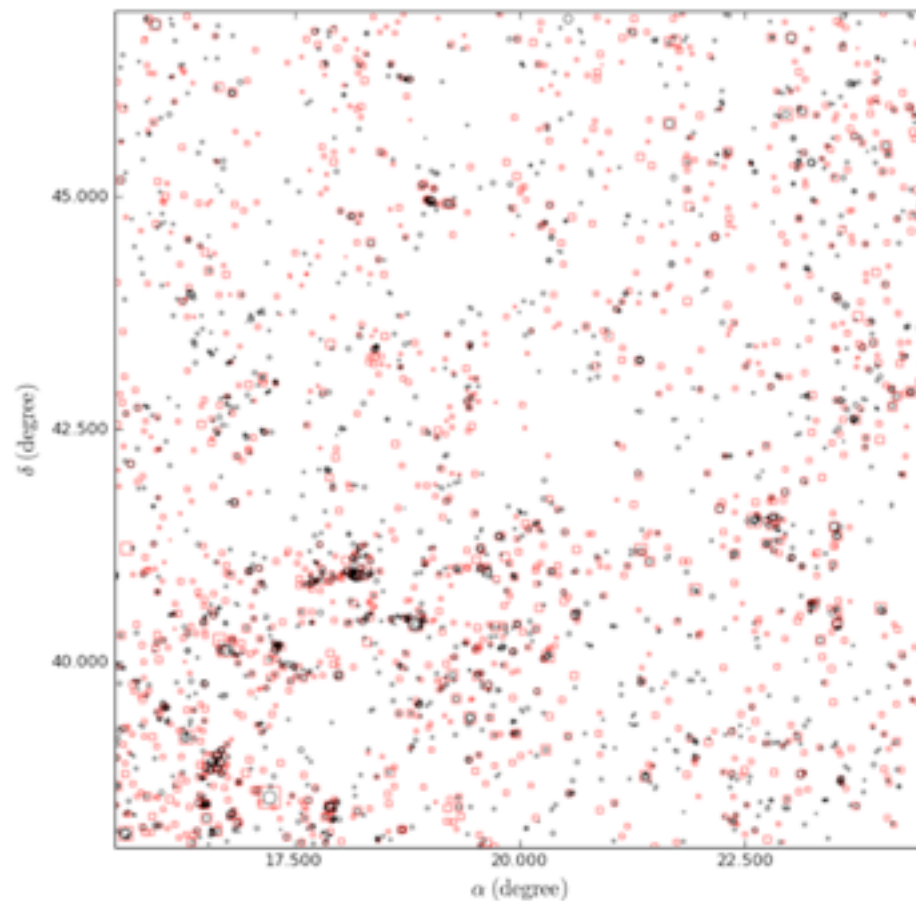
Sky Distribution $z=0.5$



Sky Distribution $z=0.6$



Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)[Clusters/Halos properties](#)[Cluster](#)

Spatial distribution of clusters (black circles) and halos (red squares, $M_{200} > 1e+13$ (M_{\odot})) in the redshift bin 0.4, scaled by their respective richness, from 0 to 50 and from 0 to 217, respectively.

Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)

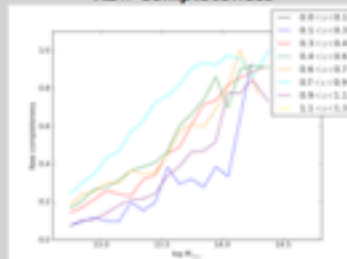
Quick Navigation

[Completeness/Purity](#)[Centering statistics](#)[Redshift statistics](#)[Mobs-Mtrue Unique](#)[Mobs-Mtrue Non-unique](#)[Cluster Comparison panels](#)[Mass-rank fit](#)

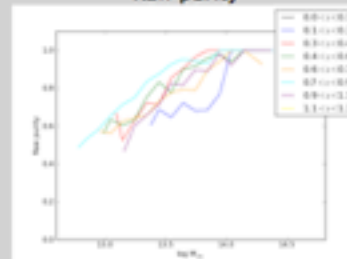
Completeness/Purity

?

Raw completeness



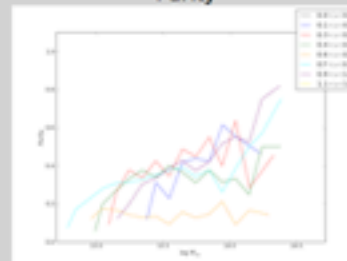
Raw purity



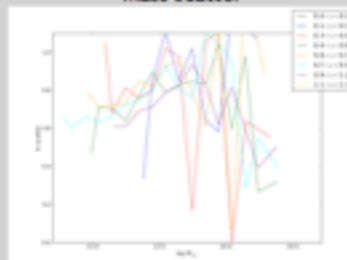
Completeness



Purity



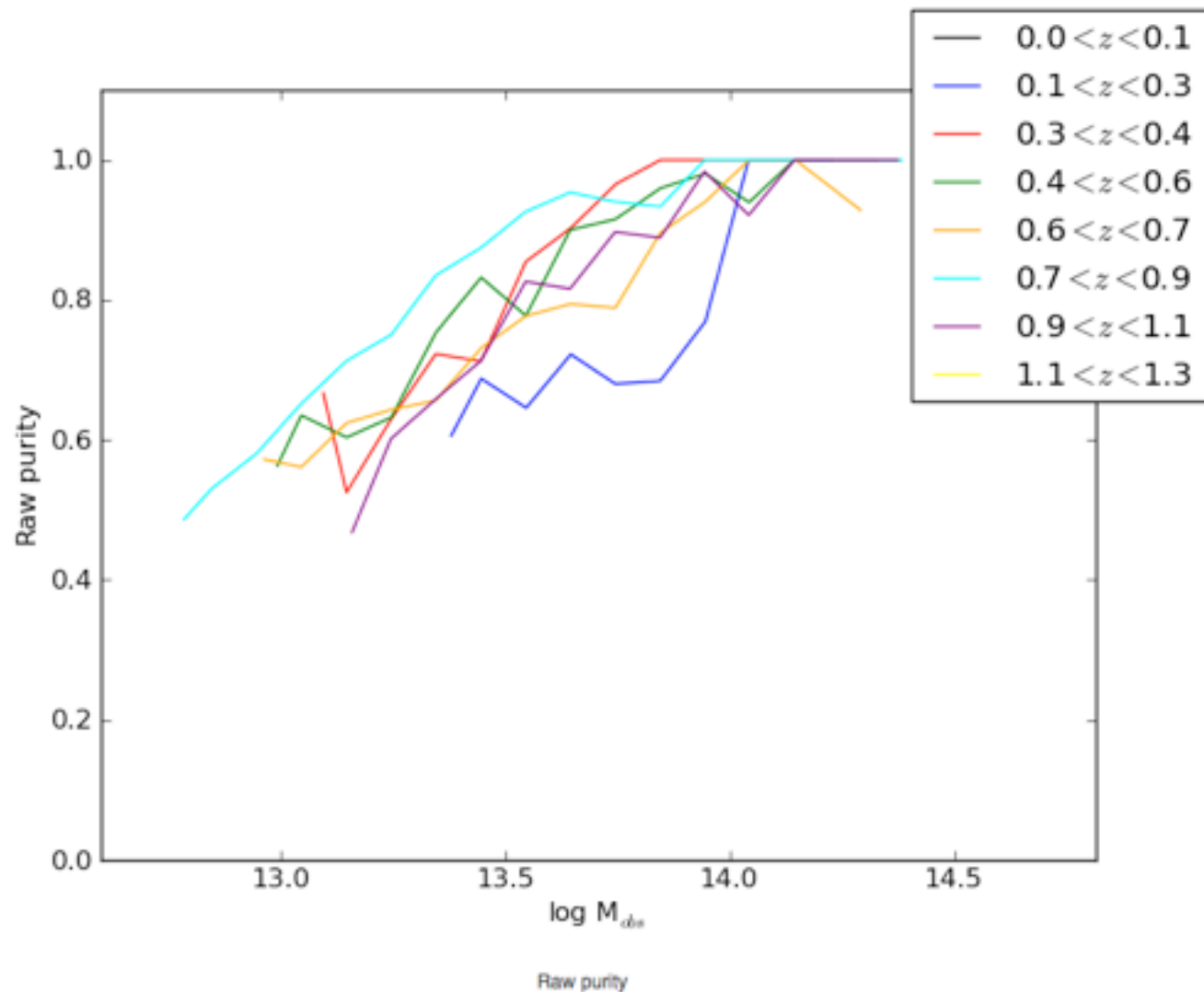
Mass scatter



Cluster Finder (WAZP)

[View Processes](#)
[Results](#)
[Comments](#)
[Summary](#)
[Data Organizer](#)
[Consolidator](#)
[Generic Plots](#)
[Cluster Comparison](#)
[QA](#)
[View unmatched](#)
[CWG tests](#)
[Science Requirements](#)
[Matching metrics](#)

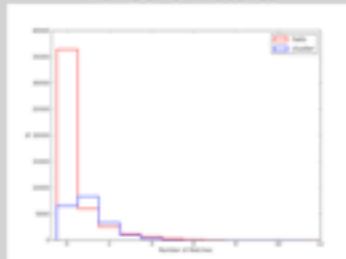
Quick Navig

[Completeness/Purity](#)
[Centering statistics](#)
[Redshift statistics](#)
[Mobs-Mtrue Unique](#)
[Mobs-Mtrue Non-unit](#)
[Cluster Comparison](#)
[Mass-rank fit](#)


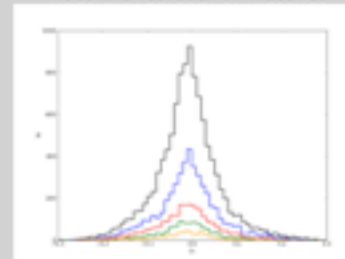
Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)[Matching](#)

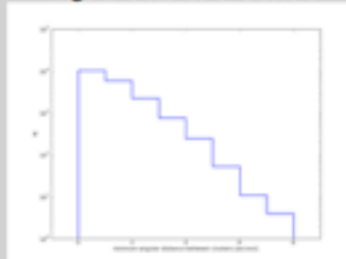
Number of matches



Redshift residual of matches



Angular distance of clusters



Cluster Finder (WAZP)

[View Processes](#)
[Results](#)
[Comments](#)
[Summary](#)
[Data Organizer](#)
[Consolidator](#)
[Generic Plots](#)
[Cluster Comparison](#)
[QA](#)
[View unmatched](#)
[CWG tests](#)
[Science Requirements](#)
[Matching metrics](#)
[Unmatched Clusters](#)
[Oddly Matched Clusters](#)

Quick Navigation

[Cluster Rank: 8](#)
[Cluster](#)
[Nearby Halos](#)
[Cluster Rank: 18](#)
[Cluster](#)
[Nearby Halos](#)
[Cluster Rank: 23](#)
[Cluster](#)
[Nearby Halos](#)
[Cluster Rank: 28](#)
[Cluster](#)
[Nearby Halos](#)
[Cluster Rank: 35](#)
[Cluster](#)
[Nearby Halos](#)

Unmatched Clusters ?

Cluster Rank: 8 ?

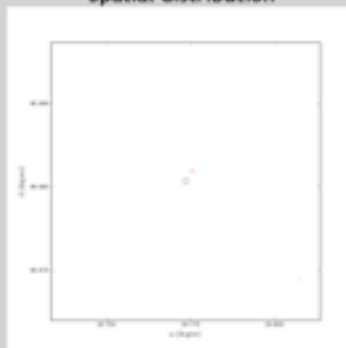
Cluster

Rank	RA	Dec	z	Ngals
8	19.773	40.480	0.932	64

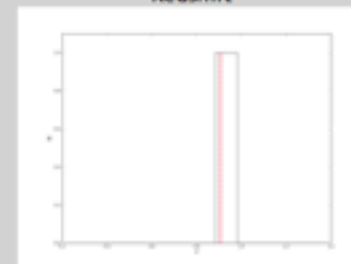
Nearby Halos

Rank	RA	Dec	z	Ngals	Mass
215	19.776	40.481	0.905	27	2.0E+14
56811	19.807	40.473	0.906	2	1.2E+13

Spatial distribution



Redshift


Cluster Rank: 18 ?

Cluster

Cluster Finder (WAZP)

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Data Organizer](#)[Consolidator](#)[Generic Plots](#)[Cluster Comparison](#)[QA](#)[View unmatched](#)[CWG tests](#)[Science Requirements](#)[Matching metrics](#)[The Sarah Hansen Chronicles](#)

Quick Navigation

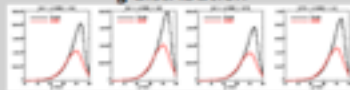
[Number counts \(!mere bogus display!\)](#)[Colors](#)[Galaxy recovery](#)[Red sequence](#)[Richness](#)

The Sarah Hansen Chronicles ?

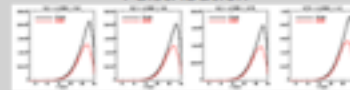
Number counts (!mere bogus display!)

?

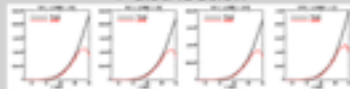
g distribution



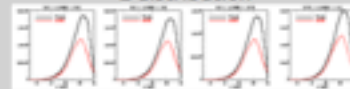
i distribution



r distribution



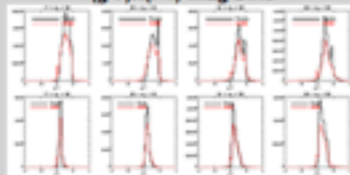
z distribution



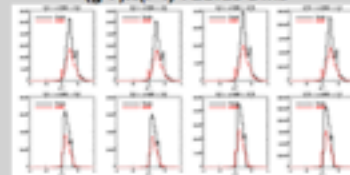
Colors

?

(g-r)x(r-i) mag bins



(g-r)x(r-i) radius bins



Galaxy recovery

?

Galaxy recovery flagmax=0



Galaxy recovery flagmax=3



Cluster Finder (WAZP)

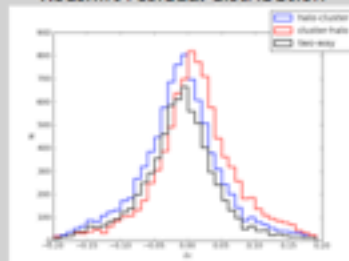
[View Processes](#)
[Results](#)
[Comments](#)
[Summary](#)
[Data Organizer](#)
[Consolidator](#)
[Generic Plots](#)
[Cluster Comparison](#)
[QA](#)
[View unmatched](#)
[CWG tests](#)
[Science Requirements](#)
[Matching metrics](#)

Quick Navigation

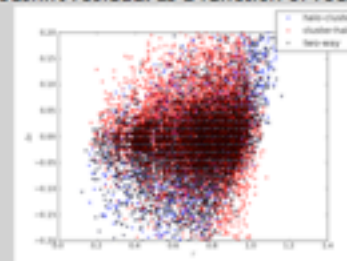
[R24: redshift statistics](#)
[R32: mass richness scatter](#)

R24: redshift statistics ?

Redshift residual distribution



Redshift residual as a function of redshift



Science Requirement 24 status

Uncertainty in the bias of cluster photo-z $\sigma(z_{\text{bias}}) < 0.03$

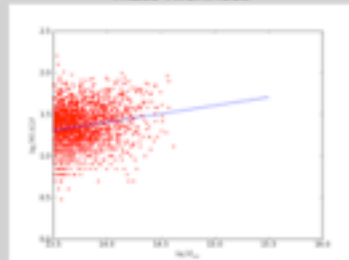
0.0007 ✓

Uncertainty in dispersion of cluster photo-z $\sigma(\sigma_z) < 0.0016$

0.00006084 ✓

R32: mass richness scatter ?

Mass Richness



Science Requirement 32: Mass Richness scatter

0.237

Pipeline Consolidator

View Processes Results Comments

?

User: Ricardo Ogando
 GMBG has only the search radius around a BCG as tunable parameter. ANNZ and all photozs are calculated in SLAC and included in the catalog. I will update the lengthy titles, just wanted to share the good stuff first.
2012-03-23 10:26:06

User: Luiz da Costa
 Is the ANNZ run done by us ? or is this a different catalog ?
2012-03-23 09:45:09

User: Luiz da Costa
 GMBG has no explicit magnitude limit cut ? no threshold ? we need to make these parameters explicit --- I still think that the titles of the tables should be shorter and the table caption should be placed in the help
2012-03-23 09:41:18

User: Ricardo Ogando
 The results are a bit surprising. Knowing that ANNZ, which GMBG is using, has worst performance for this version of Addgals than other photozs, GMBG, nevertheless, has better redshift performance. Another surprise is that WAZP has better centering performance. Since GMBG uses BCGs to center the cluster, and the mock catalog, by construction, is centered also on BCGs, this is quite remarkable.
2012-03-23 08:04:03

User: Ricardo Ogando
 This includes reorganization of Summary and a new tab about centering and redshift offset. In order to compare the runs besides the regular histogram, I also present the normalized histogram, since the number of detections can vary considerably.
2012-03-23 08:02:47

User: Ricardo Ogando
 Cluster Finder pipeline Consolidator update.
2012-03-23 07:55:31

Release Date: Mon Jan 9 2012 - 14:17:59 [ogando portal]



Jobs can be **shared** with groups or individuals, which can comment the results



The Dark Energy Survey

Ricardo Ogando
Welcome to the Ogando portal.

Home Release Notes My Workspace Pipelines Tools Data Server Help Credits Other Portals Logout

Science Portal

My Processes My Comments My Space My Uploads My Favorites

Filter: Completed

Shared Process: 10008350

Groups Users

Groups with permission:
No group added

Add Permission in group: Administrators

Close

Process ID	Start Time	Algorithm	Status	Saved	Shared
<input checked="" type="checkbox"/> 10008350	2012-03-26 06:22:15	---	✓	---	---
<input type="checkbox"/> 10008339	2012-03-25 08:59:59	GMBCG	✓	---	---
<input type="checkbox"/> 10008304	2012-03-23 16:52:12	---	✓	---	---
<input type="checkbox"/> 10008290	2012-03-23 07:52:50	---	✓	---	✓
<input type="checkbox"/> 10008143	2012-03-14 18:25:18	WAZP	✓	✓	✓
<input type="checkbox"/> 10008142	2012-03-14 18:17:32	WAZP	✓	✓	✓
<input type="checkbox"/> 10008141	2012-03-14 18:15:34	WAZP	✓	✓	✓
<input type="checkbox"/> 10008122	2012-03-06 17:21:33	---	✓	---	---
<input type="checkbox"/> 10008119	2012-03-06 17:07:16	---	✓	---	---
<input type="checkbox"/> 10008115	2012-03-05 18:06:16	WAZP	✓	---	---

Export Purge Save Share

The Dark Energy Survey

Ricardo Ogando
Welcome to the Ogando portal.

Science Portal

Home Release Notes My Workspace Pipelines Tools Data Server Documentation Help Credits Other Portals Logout




My Workspace


My Processes **My Comments** My Space My Uploads My Export My Configurations My Profile

Filter: My Comments

Show 10 entries

First Previous Next Last

Process ID	Start Time	End Time	Total Time	Pipeline	Algorithm	Status	Comments
10008290	2012-03-23 07:52:50	2012-03-23 07:53:10	0:00:20	Pipeline Consolidator	---	✓	
10008143	2012-03-14 18:25:18	2012-03-14 20:26:02	2:00:44	Cluster Finder	WAZP	✓	
10008142	2012-03-14 18:17:32	2012-03-14 19:13:25	0:55:53	Cluster Finder	WAZP	✓	
10008141	2012-03-14 18:15:34	2012-03-15 05:23:54	11:08:20	Cluster Finder	WAZP	✓	
10008061	2012-02-29 20:12:43	2012-03-01 07:30:42	11:17:59	Pipeline Consolidator	---	✓	
10008034	2012-02-27 18:45:43	2012-02-28 01:41:34	6:55:51	Cluster Finder	GMBCG	✓	
10008003	2012-02-26 09:36:53	2012-02-26 09:37:08	0:00:15	Pipeline Consolidator	---	✓	
10008000	2012-02-25 15:00:57	2012-02-26 18:56:55	1 day, 3:55:58	Cluster Finder	WAZP	✓	
10007953	2012-02-17 12:43:34	2012-02-17 12:43:46	0:00:12	Pipeline Consolidator	---	✓	
10007952	2012-02-17 11:47:07	2012-02-17 11:47:26	0:00:19	Pipeline Consolidator	---	✓	

✓ Ok ✗ Failed  Comment

Release Date: Mon Jan 9 2012 - 14:17:59

Copyright

Browse comments through **My Workspace**



Comments



The Dark Energy Survey

Ricardo Ogando
Welcome to the Ogando portal.

Home Release Notes My Workspace Pipelines

My Processes My Comments My Space My Uploads

Filter: My Comments

Process ID ▼	Start Time
10008290	2012-03-23 07:52:50
10008143	2012-03-14 18:25:18
10008142	2012-03-14 18:17:32
10008141	2012-03-14 18:15:34
10008061	2012-02-29 20:12:43
10008034	2012-02-27 18:45:43
10008003	2012-02-26 09:36:53
10008000	2012-02-25 15:00:57
10007953	2012-02-17 12:43:34
10007952	2012-02-17 11:47:07

Release Date: Mon Jan 9 2012 - 14:17:59

Information

User: Ricardo Ogando

Comments:

GMBCG has only the search radius around a BCG as tunable parameter. ANNZ and all photozs are calculated in SLAC and included in the catalog. I will update the lengthy titles, just wanted to share the good stuff first.

2012-03-23 10:26:06

User: Luiz da Costa

Comments:

is the ANNZ run done by us ? or is this a different catalog ?

2012-03-23 09:45:09

User: Luiz da Costa

Comments:

GMBCG has no explicit magnitude limit cut ? no threshold ? we need to make these parameters explicit

--- I still think that the titles of the tables should be shorter and the table caption should be placed in the help

2012-03-23 09:41:18

User: Ricardo Ogando

Comments:

The results are a bit surprising. Knowing that ANNZ, which GMBCG is using, has worst performance for this version of Addgals than other photozs, GMBCG, nevertheless, has better redshift performance. Another surprise is that WAZP has better centering performance. Since GMBCG uses BCGs to center the cluster, and the mock catalog, by construction, is centered also on BCGs, this is quite remarkable.

2012-03-23 08:04:03

User: Ricardo Ogando

Comments:

This includes reorganization of Summary and a new tab about centering and redshift offset.

In order to compare the runs besides the regular histogram, I also present the normalized histogram, since the number of detections can vary considerably.

2012-03-23 08:02:47

User: Ricardo Ogando

Comments:

Cluster Finder pipeline Consolidator update.

2012-03-23 07:55:31

Close

Science Portal

Logout



Algorithm	Status	Comments
---	✓	
WAZP	✓	
WAZP	✓	
WAZP	✓	
---	✓	
GMBCG	✓	
---	✓	
WAZP	✓	
---	✓	
---	✓	

Copyright



Pipeline Consolidator



- Access saved processes
- Compare side by side
 - Published metrics
 - Common output from halo-cluster matching code
- Cases:
 - Fine Tuning: WAZP
 - Cluster Challenge: WAZP VS GMBCG



WAZP Fine Tuning



- WAZP (Christophe Benoist/OCA)
 - Identification of galaxy over-densities in Ra, Dec, zphot space.
 - The underlying algorithm uses 2D and 1D density field reconstruction based on wavelet transforms.
- Tunable parameters:
 - Threshold
 - Magnitude limit
 - Parallelization (size of cells in sky partition)

Pipeline Consolidator

View Processes Results Comments

Summary Cluster Finder Consolidator Plots Centering Completeness Purity

Quick Navigation

[Data and Algorithm](#)
[Data and Algorithm Properties](#)
[Cluster Properties](#)
[Halo Properties](#)
[Matching](#)
[Mass scatter](#)
[Limits of Raw Properties](#)
[Limits of Cosmological Properties](#)

Data and Algorithm ?

Process ID	Data	Algorithm	Code version	Redshift type
10007934	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007935	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007936	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10007937	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift

Data and Algorithm Properties ?

Process ID	Magnitude limit	Total area (square degree)	Cell area (sqdeg)	Number of cells	Threshold
10007934	24.0	54.67	3.00	35	5.00
10007935	23.0	54.67	3.00	35	5.00
10007936	24.0	54.67	6.00	12	5.00
10007937	24.0	54.67	3.00	35	3.00

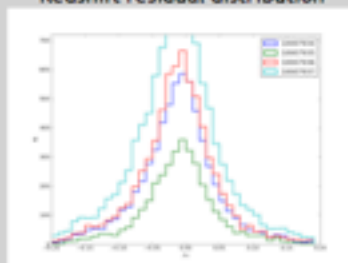
Cluster Properties ?

Process ID	Mean density	Mean redshift	Mean richness
10007934	320.2	0.73	20
10007935	164.8	0.59	17
10007936	360.2	0.73	20
10007937	681.2	0.72	17

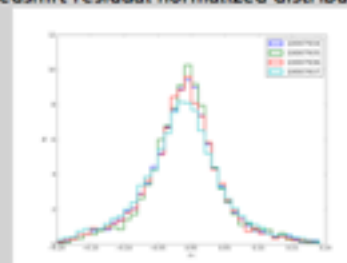
Pipeline Consolidator

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)

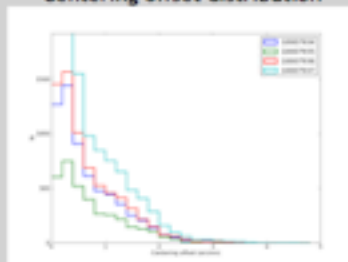
Redshift residual distribution



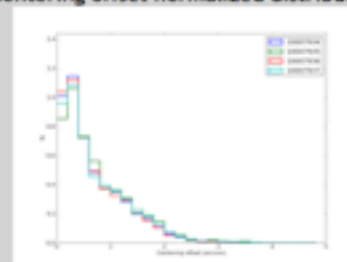
Redshift residual normalized distribution



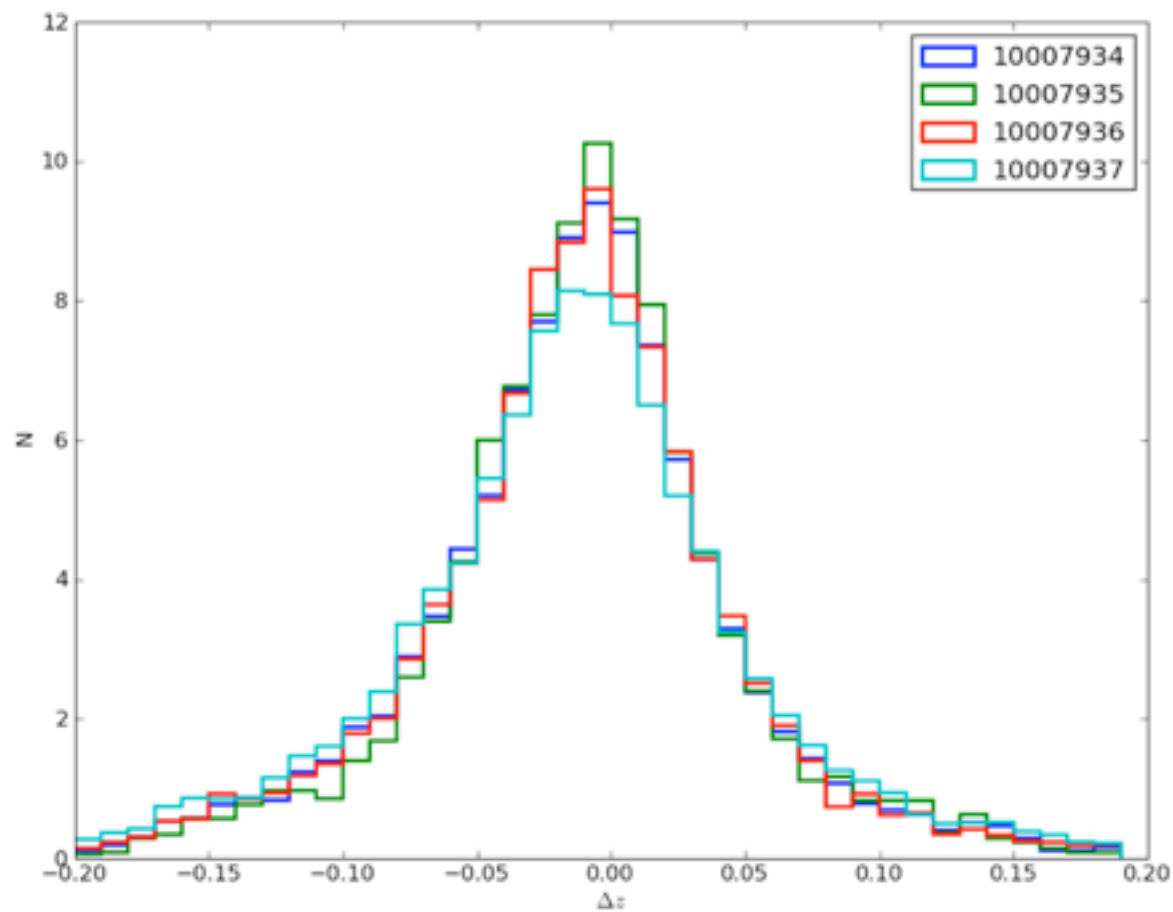
Centering offset distribution



Centering offset normalized distribution

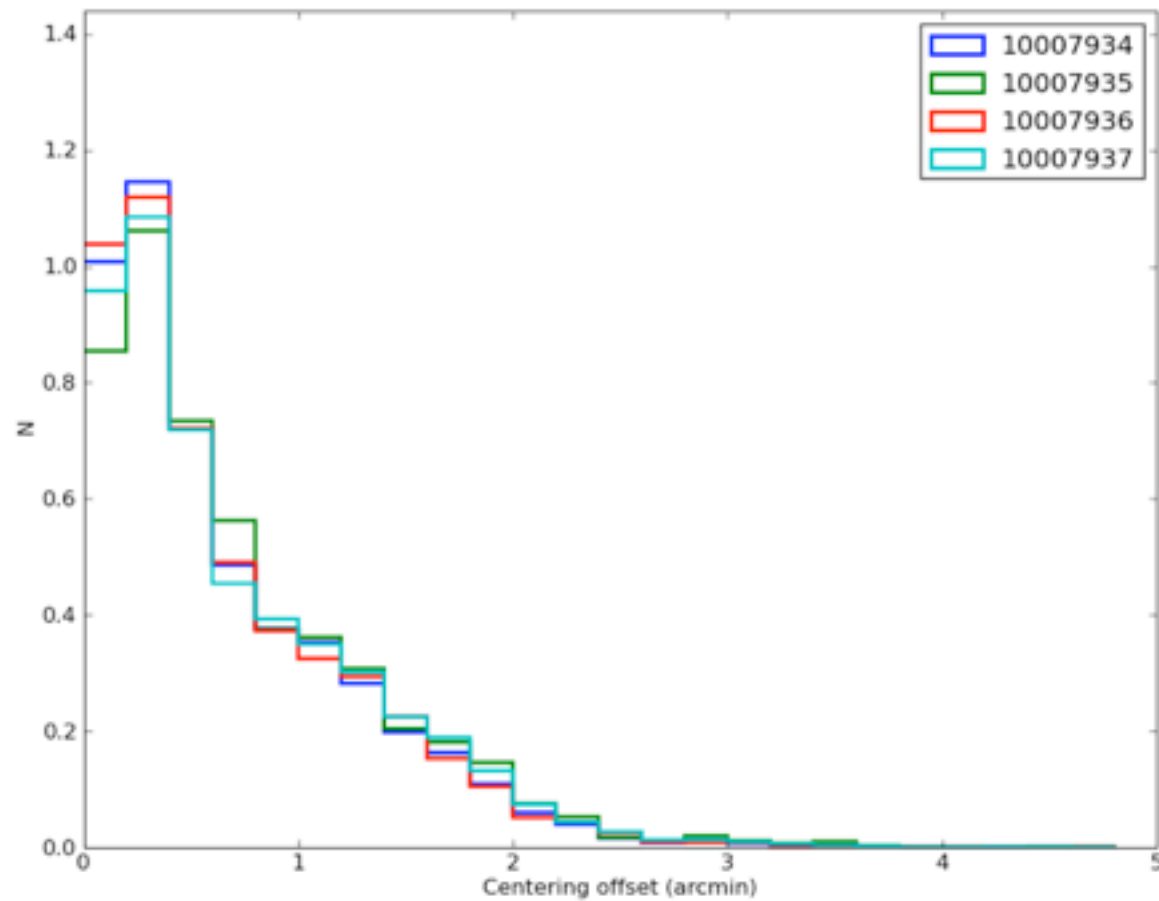


Pipeline Consolidator

[View Processes](#)**Results**[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)

Normalized distribution of redshift residual of each process.

Pipeline Consolidator

[View Processes](#)**Results**[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)

Pipeline Consolidator

[View Processes](#)
[Results](#)
[Comments](#)
[Summary](#)
[Cluster Finder Consolidator Plots](#)
[Centering](#)
[Completeness](#)
[Purity](#)
[Raw](#)
[Cosmological](#)

Quick Navigation

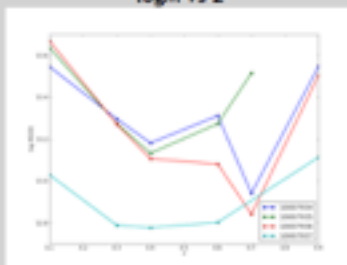
[Raw completeness 30%](#)
[Raw completeness 60%](#)
[Raw completeness 90%](#)

Raw completeness 30%

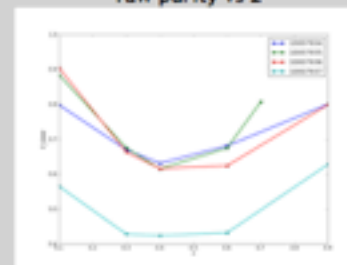
?

Raw ?

logM vs z



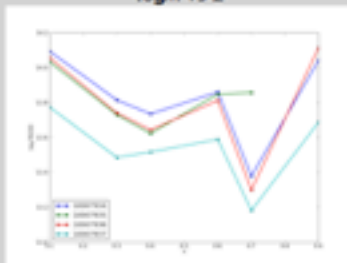
raw purity vs z



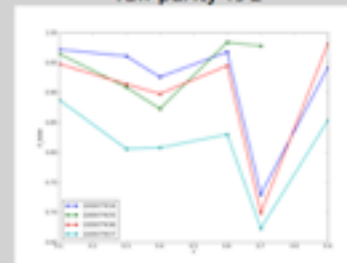
Raw completeness 60%

?

logM vs z



raw purity vs z



Raw completeness 90%

?

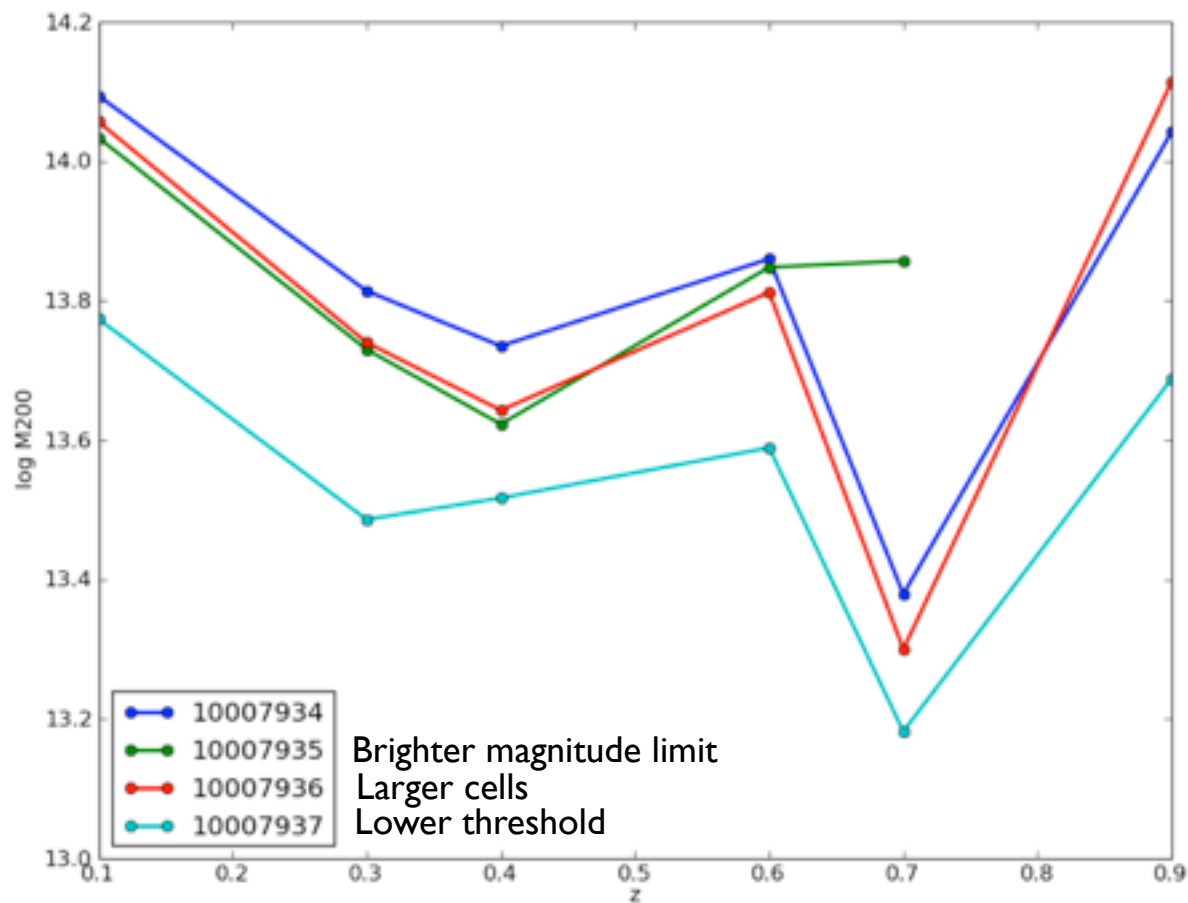
logM vs z

raw purity vs z

Pipeline Consolidator

[View Processes](#)**Results**[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)[Raw](#)[Cosmo](#)

Quick Navigation

[Raw completeness 30%](#)[Raw completeness 60%](#)[Raw completeness 90%](#)

Plot of log(Mass) vs redshift for different processes at raw completeness=0.6



Cluster Challenge



- WAZP VS GMBCG
- GMBCG searches for over-densities of red sequence galaxies + BCG. It uses photometric redshifts to assign the cluster's redshift.

Pipeline Consolidator

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)

Quick Navigation

[Data and Algorithm](#)[Data and Algorithm Properties](#)[Cluster Properties](#)[Halo Properties](#)[Matching](#)[Mass scatter](#)[Limits of Raw Properties](#)[Limits of Cosmological Properties](#)

Data and Algorithm

Process ID	Data	Algorithm	Code version	Redshift type
10008000	Addgals v3.04	WAZP	wazp-0.3-3	Gaussian error photometric redshift
10008034	Addgals v3.04	GMBG	gmbcg-41-1:+	ANNZ

Data and Algorithm Properties

Process ID	Magnitude limit	Total area (square degree)	Area of individual cell (square degree)	Number of cells	Threshold
10008000	24.0	202.10	6.00	48	3.00
10008034	---	202.10	50.00	4	---

Cluster Properties

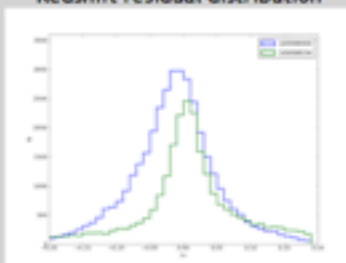
Process ID	Cluster Mean density	Cluster Mean redshift	Cluster Mean richness
10008000	657.5	0.72	16
10008034	468.7	0.66	8

Halo Properties

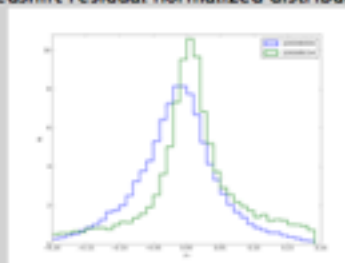
Pipeline Consolidator

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)

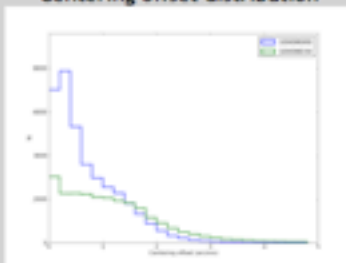
Redshift residual distribution



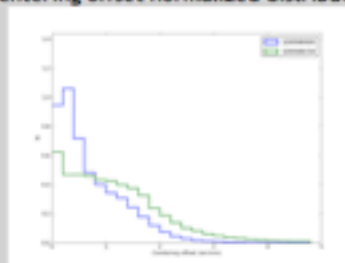
Redshift residual normalized distribution



Centering offset distribution



Centering offset normalized distribution



Pipeline Consolidator

View Processes

Results

Comments

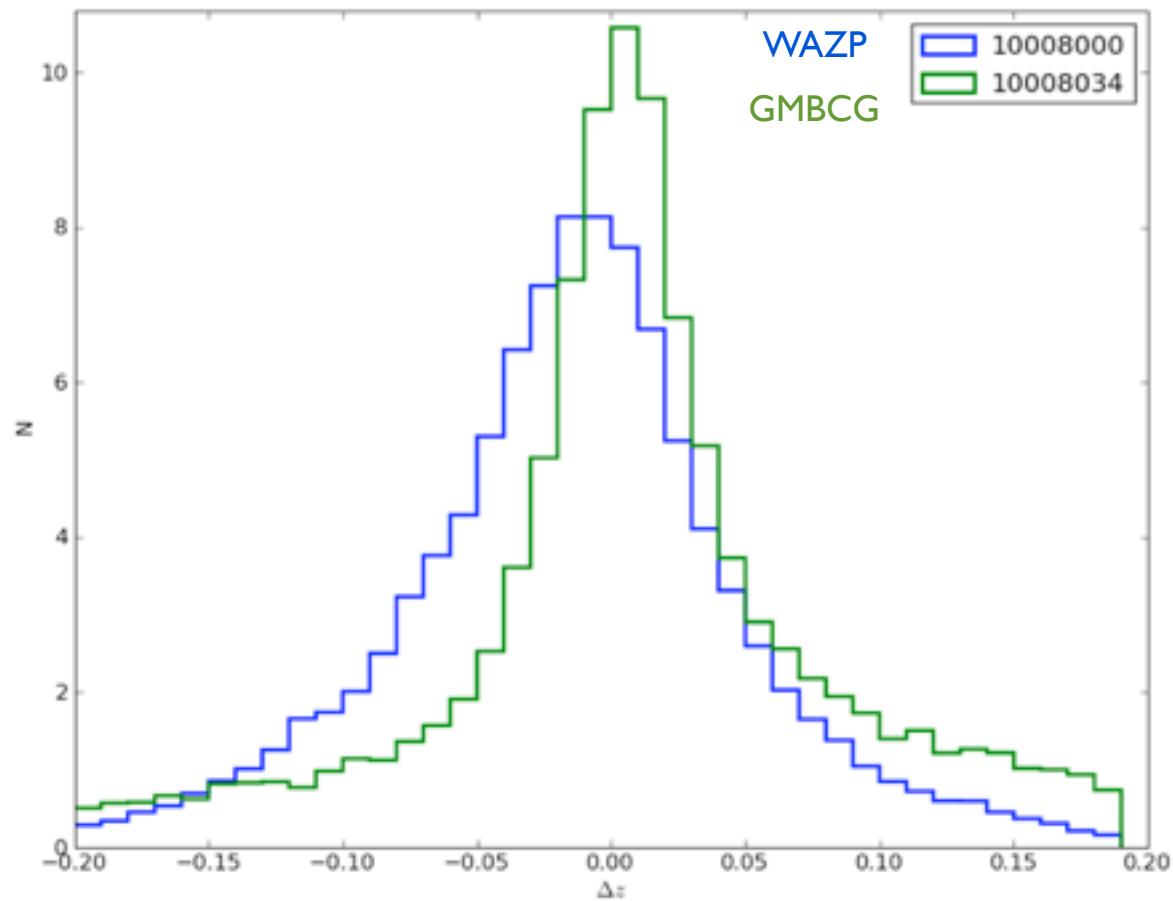
Summary

Cluster Finder Consolidator Plots

Centering

Completeness

Purity



Normalized distribution of redshift residual of each process.

Pipeline Consolidator

View Processes

Results

Comments

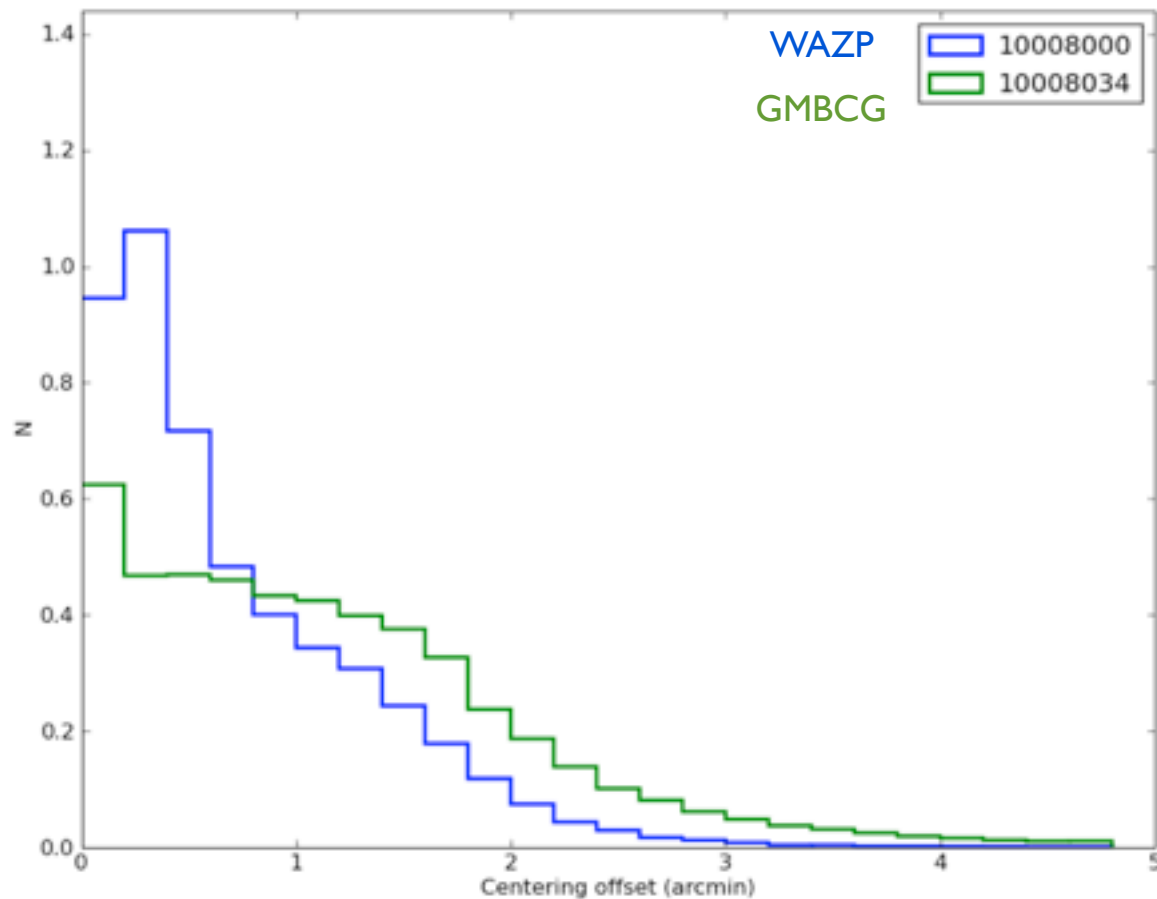
Summary

Cluster Finder Consolidator Plots

Centering

Completeness

Purity



Centering offset normalized distribution of each process.

Pipeline Consolidator

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)[Raw](#)[Cosmological](#)

Quick Navigation

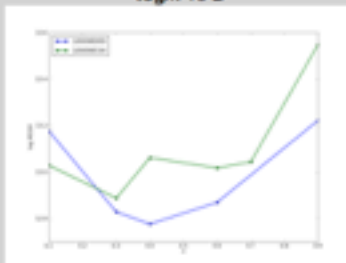
[Raw completeness 30%](#)[Raw completeness 60%](#)[Raw completeness 90%](#)

Raw completeness 30%

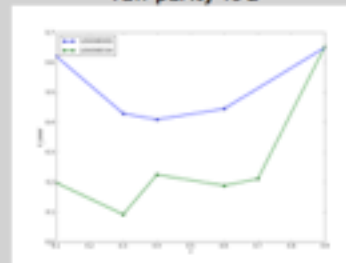
?

Raw ?

logM vs z



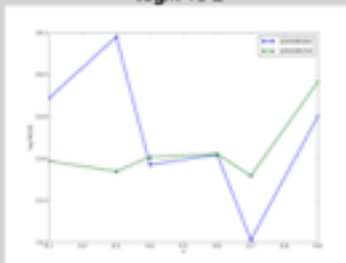
raw purity vs z



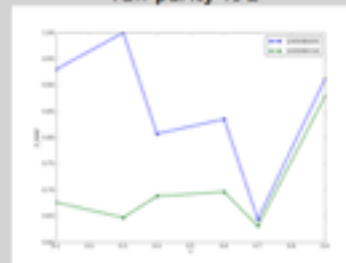
Raw completeness 60%

?

logM vs z



raw purity vs z



Raw completeness 90%

?

logM vs z

raw purity vs z

Pipeline Consolidator

[View Processes](#)[Results](#)[Comments](#)[Summary](#)[Cluster Finder Consolidator Plots](#)[Centering](#)[Completeness](#)[Purity](#)[Raw](#)[Cost](#)[Quick Nav](#)[Raw completeness](#)[Raw completeness](#)[Raw completeness](#)

Plot of $\log(\text{Mass})$ vs redshift for different processes at raw completeness=0.6

PHOTO-Z

Pipeline Consolidator

View Processes Results Comments

Summary Photo-z Compute Consolidator Plots

Quick Navigation

[Algorithm](#)
[Configuration](#)
[General statistics](#)
[Metrics](#)

Algorithm ?

Process ID	Algorithm	Version
10008462	LePHARE	lephare-0.1-2
10008463	LePHARE	lephare-0.1-2

Configuration ?

Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics ?

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61

PHOTO-Z

Pipeline Consolidator

View Processes Results Comments

Summary Photo-z Compute Consolidator Plots

Quick Navigation

[Algorithm](#)
[Configuration](#)
[General statistics](#)
[Metrics](#)

Algorithm ?

Process ID

[10008462](#)
[10008463](#)

Algorithm

LePHARE
LePHARE

Version

lephare-0.1-2
lephare-0.1-2

Configuration ?

Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics ?

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61

PHOTO-Z

Pipeline Consolidator

View Processes Results Comments

Summary Photo-z Compute Consolidator Plots

Quick Navigation

[Algorithm](#)
[Configuration](#)
[General statistics](#)
[Metrics](#)

Algorithm ?

Process ID

[10008462](#)
[10008463](#)

Algorithm

LePHARE
LePHARE

Version

lephare-0.1-2
lephare-0.1-2

Configuration ?

Process ID	Hubble parameter	Density parameter (matter)	Density parameter (lambda)	Maximum redshift	Step in redshift	Brightest abs. magnitude	Faintest abs. magnitude	SED set	Extinction law	N(z) Prior	Applied shifts
10008462	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	3	0.003,-0.004,-0.001,0.008,0.036
10008463	70.0	0.3	0.7	1.5	0.02	-25.0	-13.0	CWW_KINNEY	SMC_prevot.dat	0	0.0,0.0,0.0,0.0,0.0

General statistics ?

Process ID	Algorithm	Mean photo-z value	Median photo-z value
10008462	LePHARE	0.66	0.62
10008463	LePHARE	0.65	0.61



Lessons learned



- CFs are in C, C++, Fortran, IDL, Perl, Python.
- Need special libraries.
- All wrapped by python modules
- Time to complete code wrapping depends on
 - have source code running
 - length of workflow,
 - length of configuration,
 - data needs (e.g. from FITS to ASCII),
 - language, libraries installation, and
 - **understanding** code.
- Some of those are first-timers' difficulties.

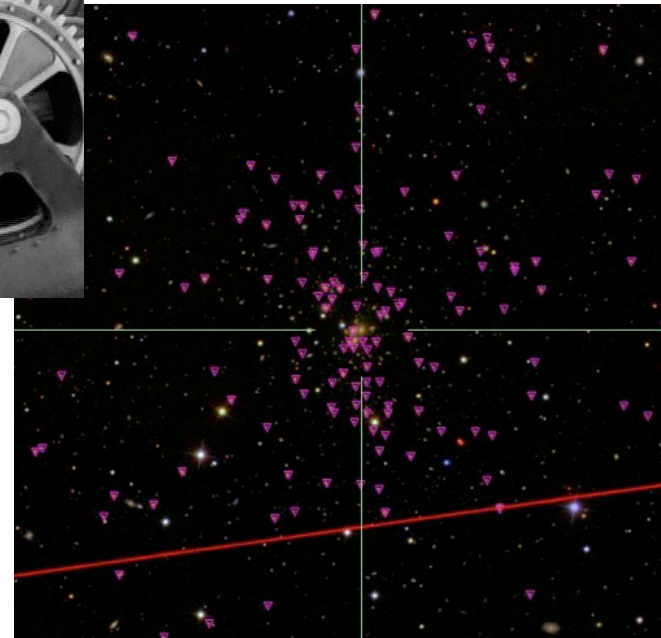
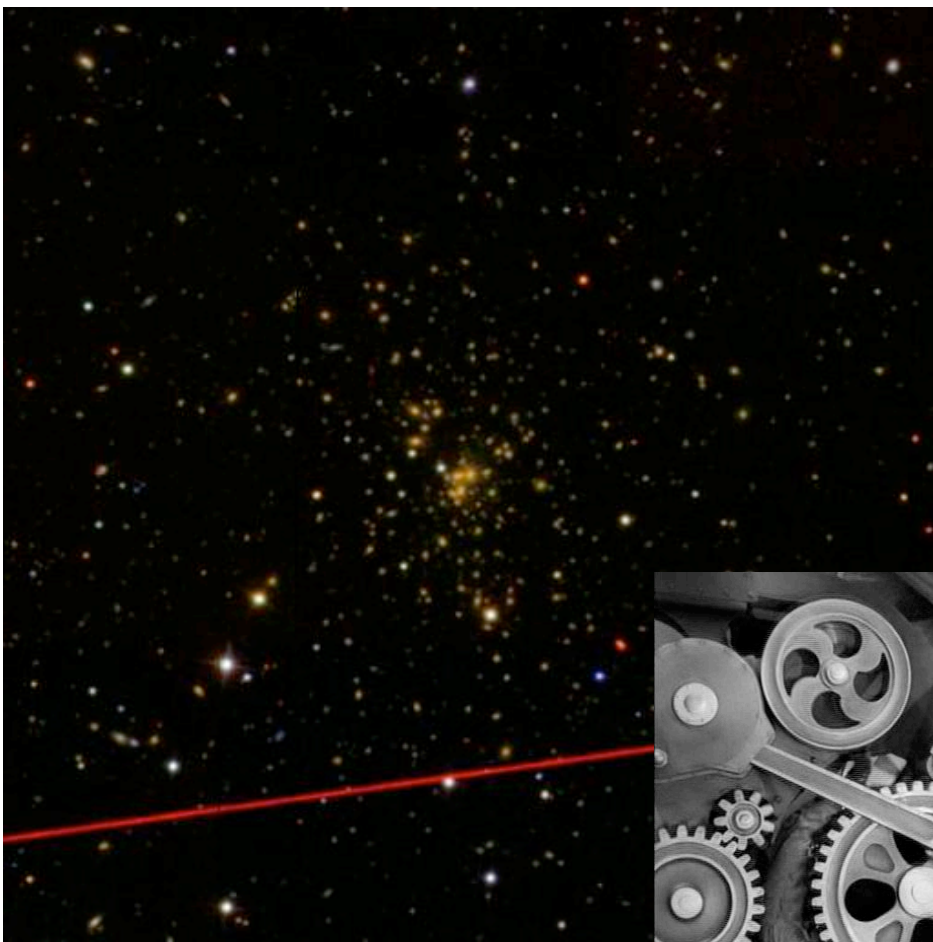
- Lots of people, codes, languages, cultures and schedules
 - Several Cluster Finders wrapped.
- Legacy. Production of cluster catalogs at the push of a button
- Pipeline Consolidator allowing easy comparison
- Future
 - close the loop with cosmology module
 - Run on real data



Thanks



Thanks

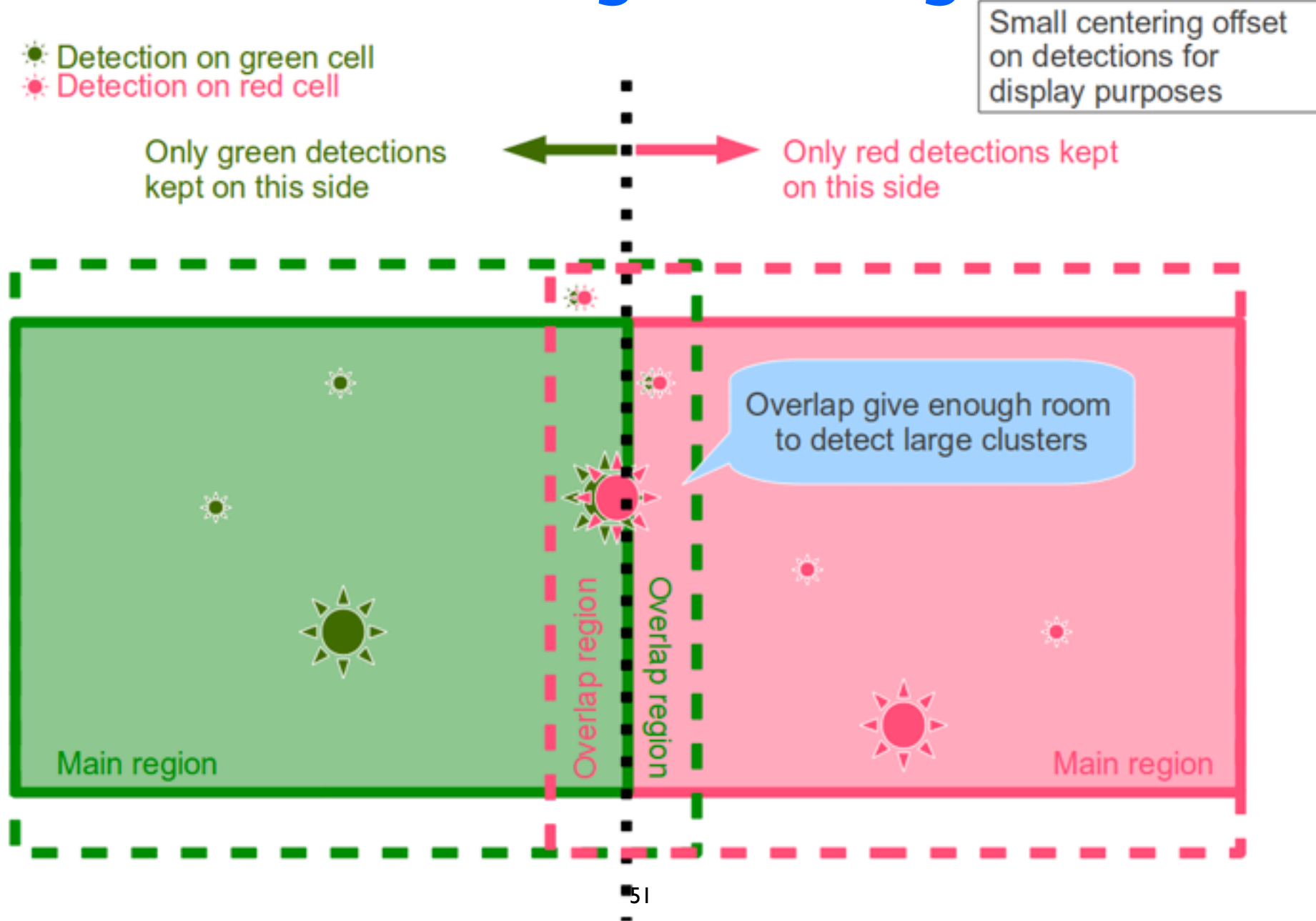




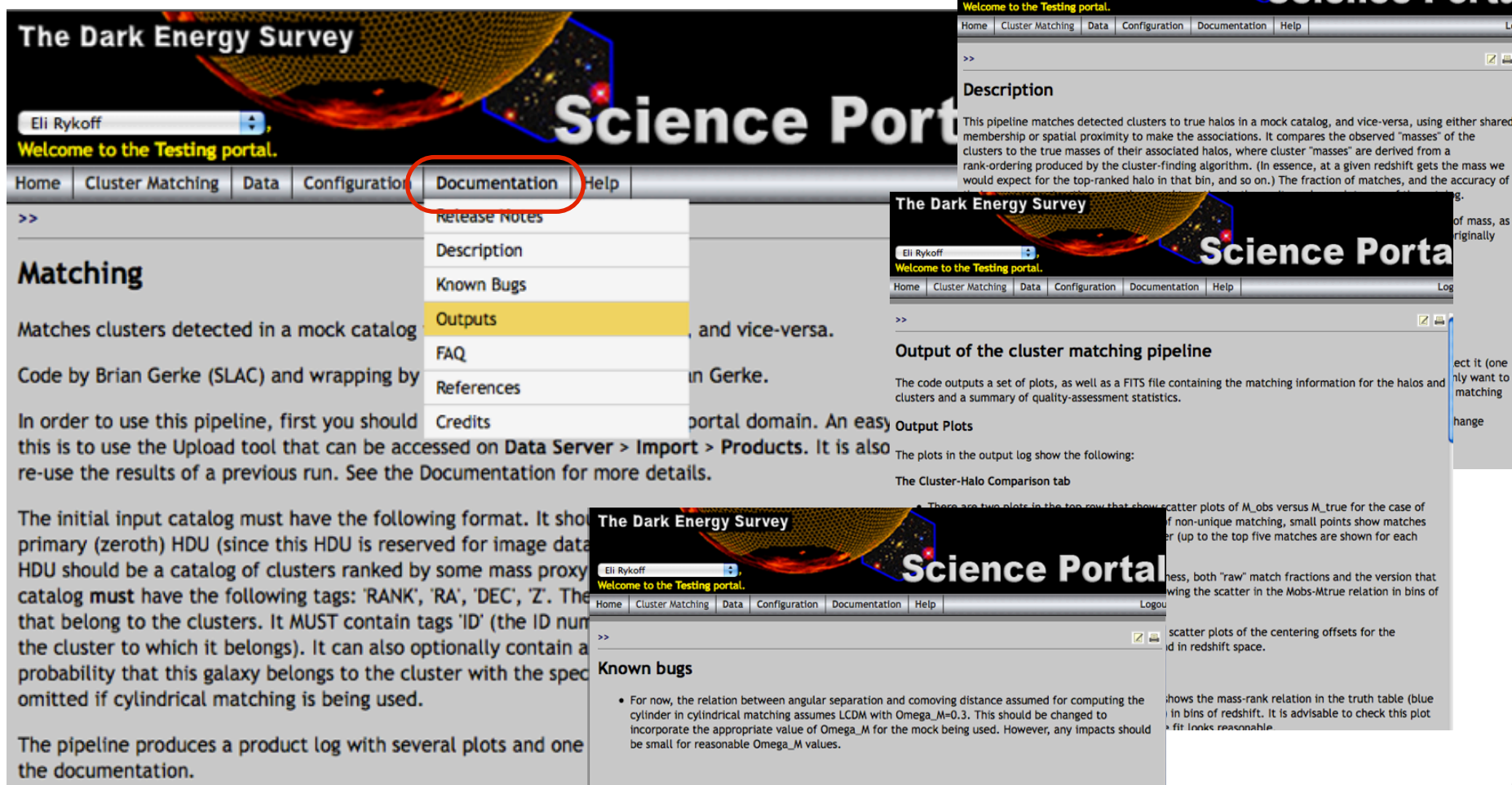
Extra Slides



Catalog stitching



Brian Gerke (SLAC) was leading this project.
He left, but the code goes on.



The screenshot displays the 'The Dark Energy Survey Science Portal' website. The user 'Eli Rykoff' is logged in, and the 'Welcome to the Testing portal.' message is visible. The navigation bar includes links for Home, Cluster Matching, Data, Configuration, Documentation, and Help. The 'Documentation' menu is open, showing options like Release Notes, Description, Known Bugs, Outputs, FAQ, References, and Credits. The 'Matching' section is highlighted, and the 'Description' section is visible on the right. The 'Known bugs' section at the bottom lists a bug related to the relation between angular separation and comoving distance.

The Dark Energy Survey Science Portal

Eli Rykoff
Welcome to the Testing portal.

Home Cluster Matching Data Configuration **Documentation** Help

>>

Matching

Matches clusters detected in a mock catalog

Code by Brian Gerke (SLAC) and wrapping by

In order to use this pipeline, first you should

this is to use the Upload tool that can be accessed on **Data Server > Import > Products**. It is also re-use the results of a previous run. See the Documentation for more details.

The initial input catalog must have the following format. It should have a primary (zeroth) HDU (since this HDU is reserved for image data). The first HDU should be a catalog of clusters ranked by some mass proxy. The catalog must have the following tags: 'RANK', 'RA', 'DEC', 'Z'. The catalog must also have tags that belong to the clusters. It MUST contain tags 'ID' (the ID number of the cluster to which it belongs). It can also optionally contain a tag 'PROB' (the probability that this galaxy belongs to the cluster with the specified ID). This tag is omitted if cylindrical matching is being used.

The pipeline produces a product log with several plots and one file. See the documentation.

Description

This pipeline matches detected clusters to true halos in a mock catalog, and vice-versa, using either shared membership or spatial proximity to make the associations. It compares the observed "masses" of the clusters to the true masses of their associated halos, where cluster "masses" are derived from a rank-ordering produced by the cluster-finding algorithm. (In essence, at a given redshift gets the mass we would expect for the top-ranked halo in that bin, and so on.) The fraction of matches, and the accuracy of the match, are output.

Output of the cluster matching pipeline

The code outputs a set of plots, as well as a FITS file containing the matching information for the halos and clusters and a summary of quality-assessment statistics.

Output Plots

The plots in the output log show the following:

The Cluster-Halo Comparison tab

There are two plots in the top row that show scatter plots of M_{obs} versus M_{true} for the case of non-unique matching, small points show matches that are not unique. The top five matches are shown for each cluster.

The bottom row shows scatter plots of the centering offsets for the cluster-halo matching in bins of redshift space.

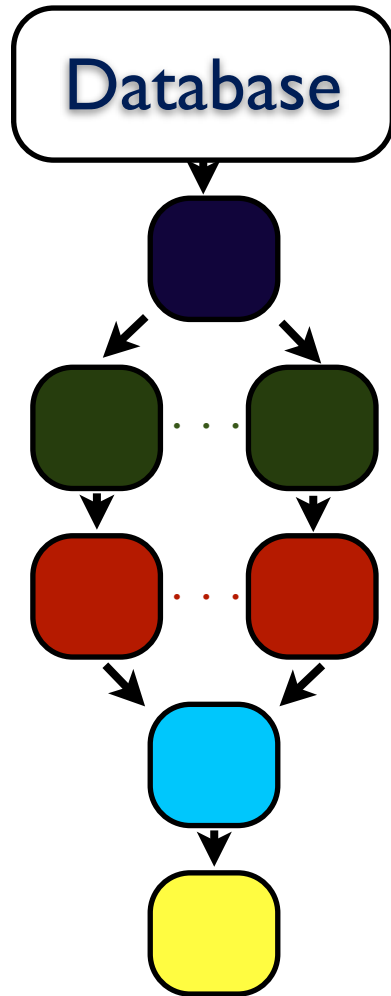
The mass-rank relation in the truth table (blue line) shows the mass-rank relation in the truth table (blue line) in bins of redshift. It is advisable to check this plot to see if the fit looks reasonable.

Known bugs

- For now, the relation between angular separation and comoving distance assumed for computing the cylinder in cylindrical matching assumes LCDM with $\Omega_{\text{M}}=0.3$. This should be changed to incorporate the appropriate value of Ω_{M} for the mock being used. However, any impacts should be small for reasonable Ω_{M} values.



Parallelization in the portal



Data Organizer: multiple queries of configurable area size **1toN**

Data Retriever: execute queries and produces FITS files **1to1**

Cluster Finder **1to1**

Consolidator: join catalogs **Nto1**

Cluster Comparison, Plots... **1to1**