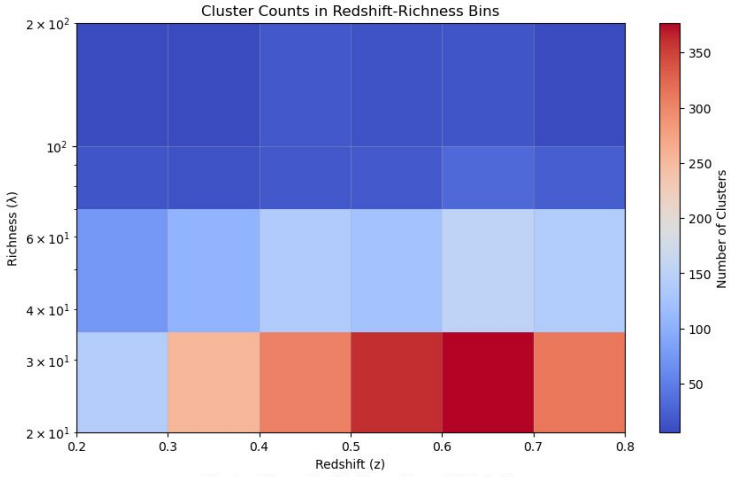
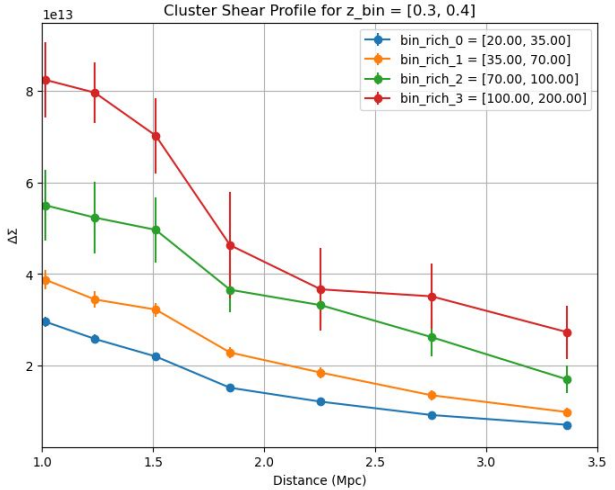
DESC Cluster Pipeline Updates: Full Chain Analysis

LSST-France 12/06 Orsay
Eduardo José Barroso
Post-doc@LAPP

Cluster Introduction

- Galaxy clusters are the largest gravitationally bound structures in the universe, and we study them to understand dark matter, cosmic evolution, and the physics of extreme environments
- Their formation trough Gravitational collapse is directly related to large structure formation → Ω_m σ_8
- We are using two different probes: The number of cluster inside a region, cluster counts, and their stacked shear profile.
- The figures are from the redMaPPer cluster catalog and we show the probes in regions of redshift and richness bins.

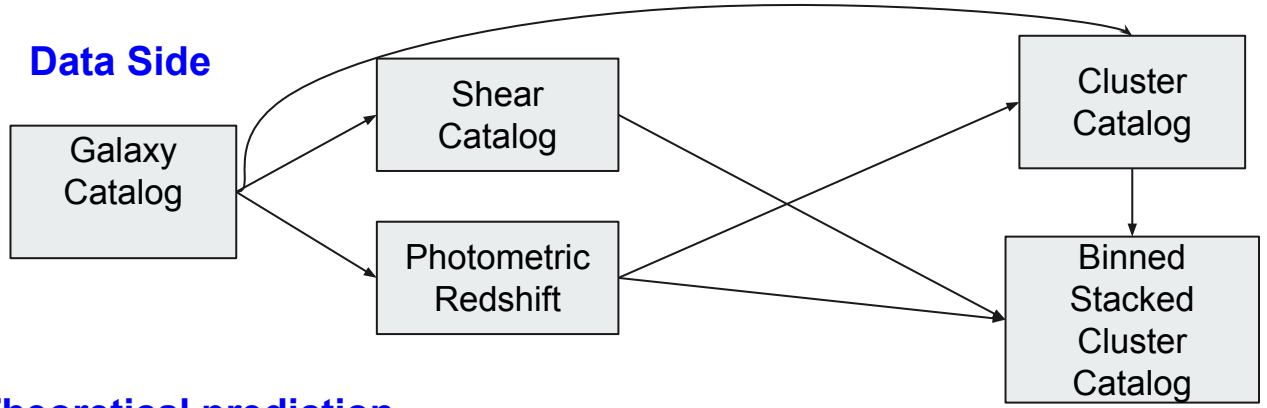




Cluster Pipeline

➔ A primary goal is to obtain constraints on Ω_m and σ_8 from the Λ - CDM

Data Side

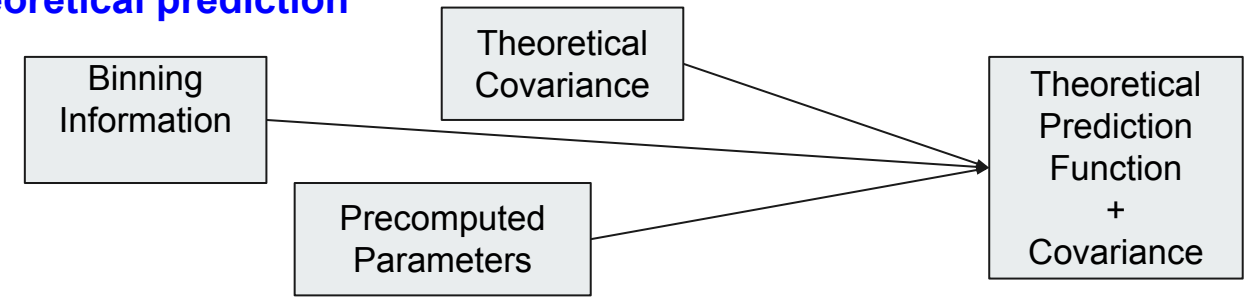


➔ We use stacked excess density surface mass instead of just the shear

➔ Theoretical covariance for counts

➔ Data driven for shear

Theoretical prediction

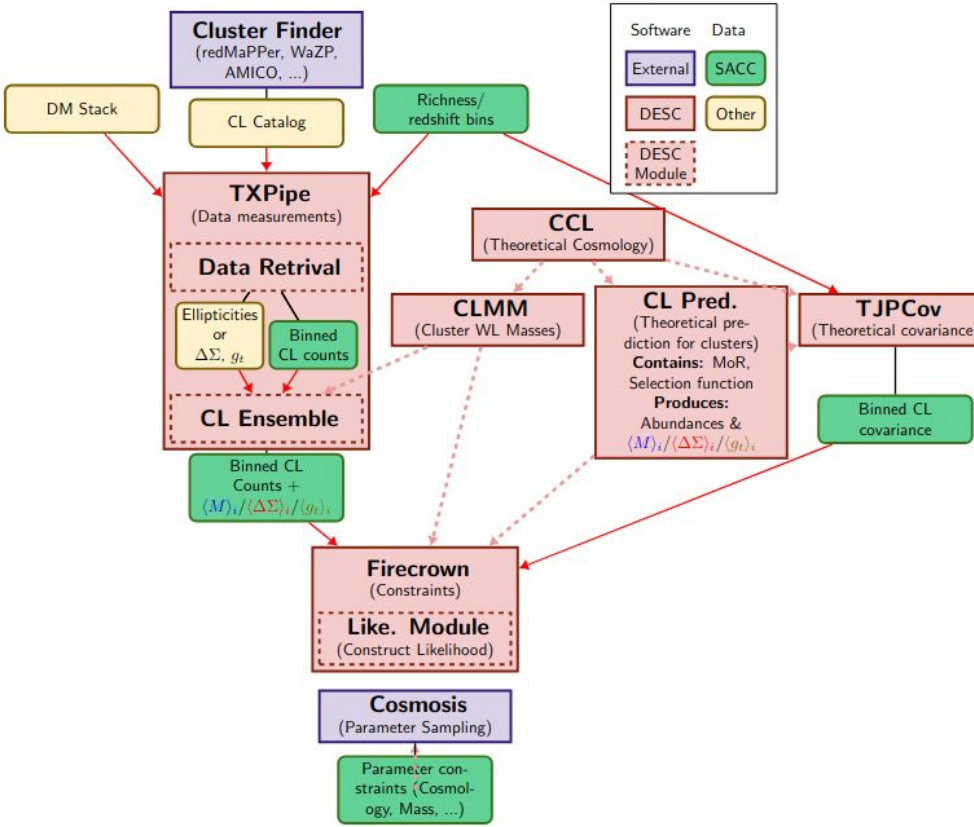


CLuster Pipeline Diagram

- ➔ The data calculations are done inside TXPipe
- ➔ The theoretical Covariance is done in TJPCov
- ➔ For the moment the CL Prediction module is inside Firecrown
- ➔ All outputs are combined in the likelihood module so we can run the chains

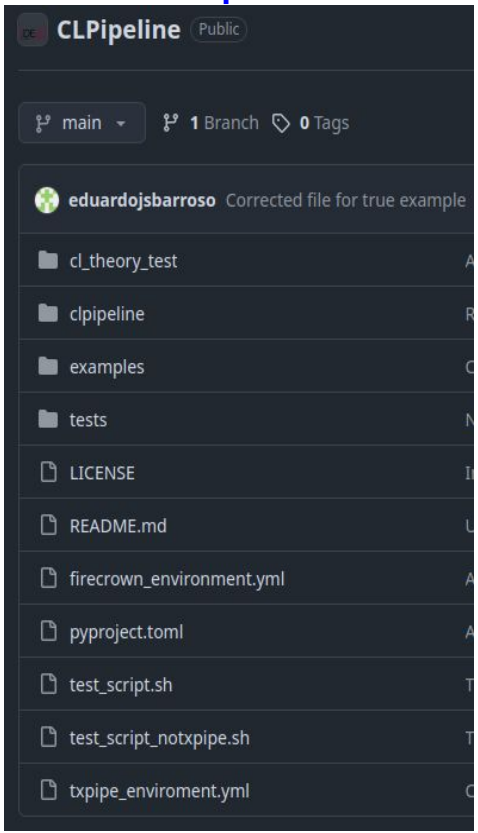
Some Recent Contributions

- ➔ Y. Zhang has been working with Firecrown and the CL Pred
- ➔ S. Elles has been investigating computing resources
- ➔ Discussions on the Pipeline with C. Payerne and C. Combet
- ➔ Cluster code merged in TXPipe from M. Ricci and J. Zuntz



CLPipeline Github repository

Repository to run all the stages of the Cluster Pipeline



Everything is divided in 3 stages:

- ➔ **TXPipe Stage**: Data computation
- ➔ **TJPCov Stage**: Covariance Calculation
- ➔ **Firecrown**: Theoretical prediction preparation and Firecrown required files for MCMC sampling

The stages are run using Ceci, a pipeline tool from LSST-DESC

The code is based on two yml files

Pipeline FIle

- ➔ Which stages are going to be run
- ➔ Inputs and outputs definitions
- ➔ Computing Performance

Configuration FIle

- ➔ Configurations specifications for each stage

CLPipeline Separated into 3 stages

TXPipe Stage

- ➔ Ingest Cluster and Galaxy catalogs
- ➔ Perform shear calibration
- ➔ Source Selector
- ➔ Source redshift estimation (rail pipeline)
- ➔ Cluster Shear Catalogs with stacked profiles (clmm)

Metacalibration or Metadetect

Perform redshift/snr and other cuts and compute selection biases

BPZ algorithm. The rail pipeline is already integrated in TXPipe

We are able to produce our own catalogs with calibrated shear and estimated photometric redshifts

TJPCov Stage

- ➔ Cluster Counts Theoretical covariance
- ➔ External code validation (Alessandra Fumagalli)

Firecrown Stage

- ➔ Define the theoretical prediction
- ➔ Fitting for: cosmology parameters, cluster concentration, MOR, etc

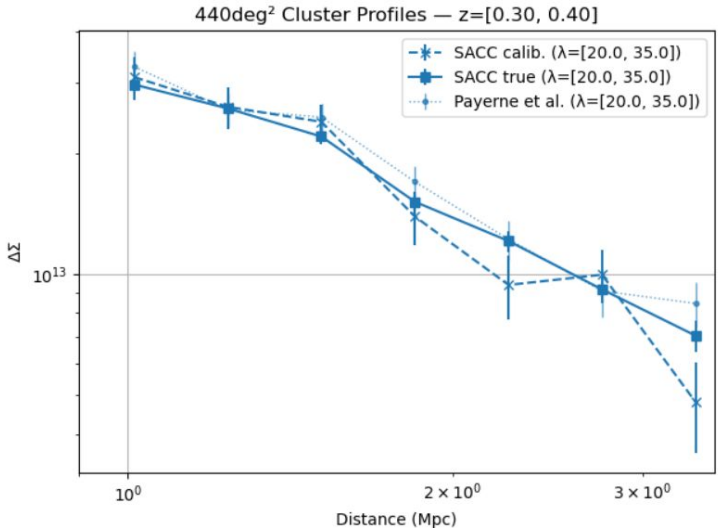
CLPipeline

- ➔ To run the Pipeline, you must submit the jobs defined on each example
- ➔ For most of the examples, the input data is saved on a shared space on either Nersc or IN2P3
- ➔ There are two predefined conda environments (recent examples done in IN2P3)

To run the examples I will show here, go to this Github and check the instructions

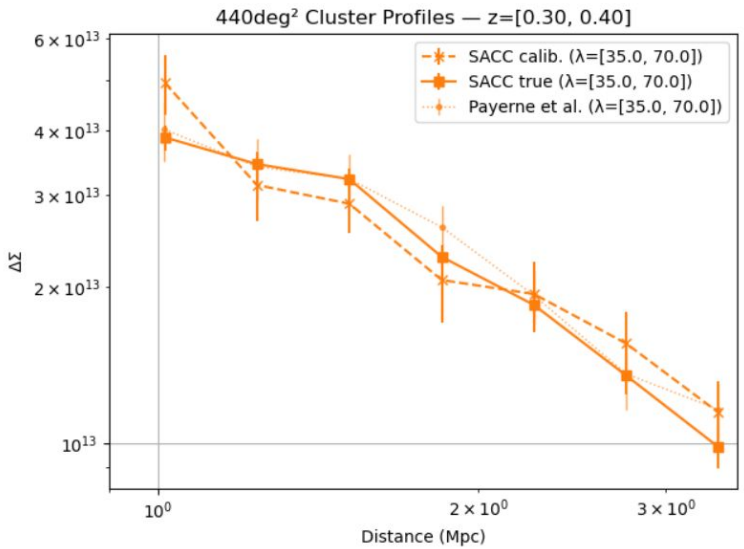
- `CL_cosmoDC2-full_concat_in2p3.yml` : This is the pipeline configuration file that sets which stages shall be run and with which configurations.
- `cosmodc2_config_in2p3.yml` : This is the configuration file for each stage run in the pipeline configuration file
- `launch_job_in2p3.sh` : This is the bash script to be submitted where we activate the right conda environments to run each stage of the pipeline.
- `launch_job_firecrown.sh` : Another script to launch the MCMC sampler.

CLPipeline full chain analysis with cosmoDC2 - Preliminary tests



Source Selection

- ➔ r < 26.9 , i<26.2 , z<25.5. Similar to expected magnitude depth for 10-Year LSST with SNR=4
- ➔ Behind: $z_{cosmoDC2} > z_{cl} + 0.2$
- ➔ $\sigma_{SN} = 0.26$



Weak Lensing Profile

- ➔ Richness and redshift bins
- ➔ R ∈ [1.0, 3.5] Mpc
- ➔ z < 0.8

CLPipeline full chain analysis with cosmoDC2 - Preliminary tests

Modeling Choices

- ➔ Despali Mass function
- ➔ Fiducial cosmoDC2 cosmology
- ➔ Selection function fitted from redmapper performance analysis
- ➔ Mass-Richness relation
- ➔ One halo regime, critical concentration-mass relation
- ➔ R ∈ [1.0, 3.5] Mpc
- ➔ z ∈ [0.2, 0.8]

Cluster Counts

$$N_{ij} = \Omega \int_{z_i}^{z_{i+1}} dz \int_{\lambda_j}^{\lambda_{j+1}} d\lambda \int_{m_{\rm min}}^{m_{\rm max}} dm \, \underbrace{\frac{dn(m,z)}{dm}}_{\text{mass function}} \frac{d^2V(z)}{dz d\Omega} \underbrace{\Phi(\lambda, m, z)}_{\text{Selection function}} \underbrace{P(\lambda|m,z)}_{\text{MOR}}$$

Constrain MOR

- ➔ Combined counts+WL (Gaussian likelihood)
- ➔ Theoretical Cov for counts
- ➔ WL profiles with data driven covariance

Shear profile

$$\Delta\Sigma_{ij}(R) = \frac{1}{N_{ij}} \int_{z_i}^{z_{i+1}} dz \int_{\lambda_j}^{\lambda_{j+1}} d\lambda \int_{m_{\rm min}}^{+\infty} dm \times$$

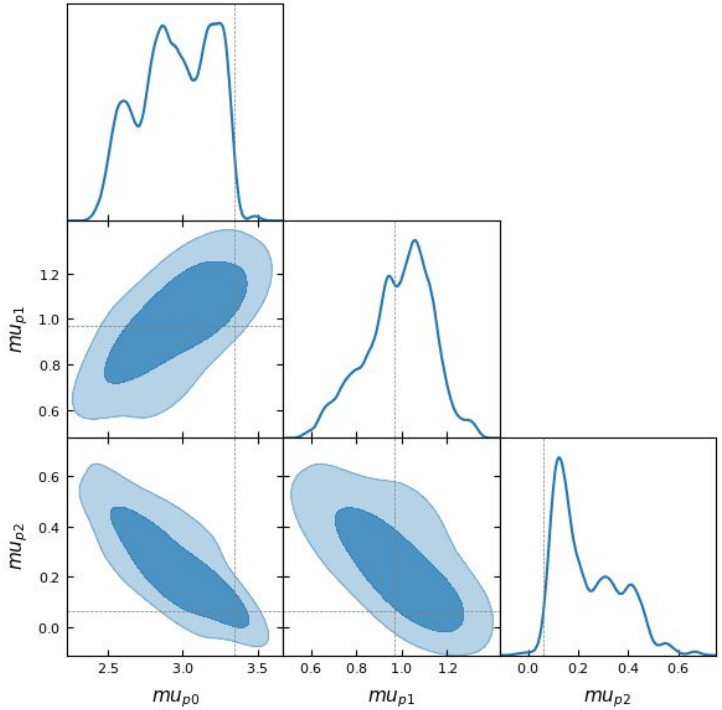
$$\frac{d^2N(m,z)}{dz dm} \underbrace{c(m,z)}_{\text{completeness}} P(\lambda|m,z) \Delta\Sigma(R|m,z).$$

Constraints on the MOR from MCMC sampling

Run parameters

- ➔ Chains have 30% burn in
- ➔ Results have 2000 samples
- ➔ Emcee sampler
- ➔ 20 walkers

<u>The fiducial values from the matched cluster and halo catalogs are within the 2σ contours</u>



Fiducial

$$\mu_{p0} = 3.35 \pm 0.01$$
$$\mu_{p1} = 0.96 \pm 0.02$$
$$\mu_{p2} = 0.06 \pm 0.08$$

Run metrics

TXPipe

- The full chain takes around 15h for the estimated redshifts and calibrated shear
  - CLClusterEnsembleProfiles stage took 9h ⟹ Local run took around 2h
  - Loading cosmoDC2 galaxy shear catalog took around 3h
  - Jobs are being run with ceci assuming from 10 to 30 mpi tasks
  - Files take around 1.4Tb (Source redshift PDF takes around 1.3Tb)

TJPCov

- Computation takes around 3 minutes
- Impossible to do it every MCMC interaction

MCMC Sampling

- Code is run only in serial
- Each likelihood interaction takes around 15 seconds
- Generated chains took around 14h

Conclusions

- **Firecrown+TXPipe+TJPCov** integration is functional.
- The code is running on a separate GitHub where everything is controlled by yml files.
- We can generate calibrated shear and photo-z catalogs within the Cluster Pipeline.
- Successfully obtained **mass–observable relation constraints** using **cluster counts** and **weak lensing** with cosmoDC2.

Next steps

- Optimize and validate computing resources
- Complexify the pipeline and prepare data
- Future plan is to validate/publish the pipeline with the DC2
- Discuss future tests with DP1 regarding the data part of the pipeline
- Contact me for feedback, insides or discussion on this repository