LSST Dark Energy Science Collaboration (DESC) and the Second Data Challenge (DC2)

Katrin Heitmann for the DESC collaboration

Argonne National Laboratory

LSST@Europe, Lyon, June 2018

This is DESC!

Collaboration Meeting

Hack week

What are we trying to learn about the Universe?

We want to understand cosmic acceleration:

- Why is the expansion of the Universe accelerating? How does the expansion work?
- What was the expansion history? How did structure grow during the expansion?
- Assuming some dark energy drives the expansion, what is its equation of state, w? And has that varied in time? Is it the same everywhere?
- Neutrinos will have played a role: how many species are there, and how much energy do they represent?
- Are we right to assume General Relativity is correct? Can we distinguish modified gravity from Dark Energy?

How do imaging surveys provide answers to these questions?

- **Weak gravitational lensing (WL)**, **galaxy clustering (LSS),** and **clusters of galaxies (CL)** all provide a direct probe of the dark matter structures on the largest scales, and how fast they grew: locate galaxies and clusters, measure galaxy shapes in images, compute correlation functions

- **Type Ia Supernovae (SN)**, **strong lens systems (SL)**, **Baryon Acoustic Oscillations** and gravitational wave sources provide standard candles, timers, rulers etc for measuring the expansion rate, independently of structure growth

DESC Forecast of Dark Energy Measurement with LSST

Forecast 68% confidence level constraints on (w_0, w_a) for individual probes and their combination after 10% (Y1, left) and all (Y10, right) LSST data is analyzed.

T. Collett, T. Eifler, E. Gawiser, R. Hlozek, R. Mandelbaum, et al.

Meeting LSST DESC Goals

To make cosmological measurements with LSST data, we need to:

- Develop, validate, maintain and operate a set of simulation, processing and analysis software pipelines and infrastructure

- Be able to re-process and analyze ~petabytes of LSST images and catalogs

- Work efficiently together to perform a complex and challenging joint inference

- Feed improved processing algorithms back to the Project/Facility as needed, validated at scale

- Science Roadmap is our guiding document (292 pages!)

LSST DESC Data Challenges

		FY16 Q1 Q2 Q3 Q4	FY17 Q1 Q2 Q3 Q4	FY18 Q1 Q2 Q3 Q4	FY19 Q1 Q2 Q3 Q4	FY20 Q1 Q2 Q3 Q4	FY21 Q1 Q2 Q3 Q4	FY22 Q1 Q2 Q3 Q4
		← LSST DESC Pre-commissioning Data Challenges →				← LSST System Commissioning		
LSST DESC Pre-commissioning Data Challenges	Data Challenge 1 (DC1)	RQ Production →	Analysis →					
	Data Challenge 2 (DC2)		RQ →	Production Analysis →				
	Data Challenge 3 (DC3)				RQ →	Production Analysis →		
LSST DESC Commissioning	ComCam Data Challenge SV Data Challenge						ComCam Analysis	SV Analysis
LSST Facility Commissioning	Early Commissioning, ComCam LSSTCam Commissioning					I&T Obs / I&T Obs		

- DESC is carrying out large-scale data challenges (DCs) to prepare for the arrival of LSST data
- Full end-to-end simulation and processing to enable
  - Testing and developing of analysis and processing pipelines
  - Testing and developing of mitigation strategies for systematic effects (e.g., sensor defects, blending, atmospheric effects, …)
  - Testing and developing data access strategies
- Three DCs with increasing complexity and size, currently data generation for DC2

Focus here: Data Challenge 2

- **Contributions from a LOT of people** (listed each name ones…)!
- Design: Humna Awan, Rahul Biswas, Eric Gawiser, Phil Marshall et al.
- Catalog production: Andrew Benson, Nick Frontiere, Salman Habib, Andrew Hearin, Katrin Heitmann, Joe Hollowed, Dan Korytov, Eve Kovacs, Patricia Larsen, Nan Li, Adrian Pope, Steve Rangel, Jim Chiang, Scott Daniels, Tom Glanzman, Danny Goldstein, Mike Jarvis, Bryce Kalmbach, Josh Meyers, James Perry, John Peterson, Glenn Sembrowski, Chris Walter, Dominique Boutigny, Nicolas Chortard, Dominique Fouchez, Fabio Hernandez, Heather Kelly, Simon Krughoff et al.
- Validation: Anita Bahmanya, Duncan Campbell, Joe deRose, Seth Diegel, Francois Lanusse, Rachel Mandelbaum, Yao Yuan Mao, Chris Morrison, Jeff Newman, Daniel Perrefort, Eli Rykoff, Javier Sanchez, Melanie Simet, Chun Hao To, Tom Uram, Vinu Vikraman, Alex Drlica Wagner, Michael Wood-Vasey, Rongpu Zhou et al.

50+ DESC Members!

DC2: Projects

- So far: more than 20 DESC Projects proposed, to use the DC2 data

- Analysis development, *system tests*

- Spanning all working groups

Concepts for DC2 Projects

📄 DC2 Project: 3x2pt analysis with DC2 catalogs — Example weak lensing + galaxy clustering analysis, going from DC2 catalogs, to 3x2pt functions, to cosmological parameters

📄 DC2 Project: Cluster Mass Calibration (CLMassMod) — Mass estimate calibration for galaxy clusters

📄 DC2 Project: Cluster Shear Calibration + Mass Estimate — Shear calibration and mass estimate for galaxy clusters

📄 DC2 Project: Comparison of Machine Learning Algorithms in Classifying Detections (Not yet) —
Use multiple algorithms that are available and study possible variations on the algorithms to improve the classification of detections into astrophysical sources and image processing artifacts

📄 DC2 Project: CosmoDC2 Production — Production of extra-galactic catalog cosmoDC2 with area of 5000 sq. deg.

📄 DC2 Project: Developing and testing DM catalog emulators — The goal of this project is to develop and test approaches to emulating DM catalogs without generating image simulations, using the DC2 simulations as a training dataset.

📄 DC2 Project: Mass-mapping from DC2 catalogs — Validate mass-mapping pipeline on DC2 data processed through DM + shape measurement pipeline

📄 DC2 Project: Non-linear bias and assembly bias modelling — Testing non-linear bias models and the effects of assembly bias using clustering information from DC2 catalogs

📄 DC2 Project: Production and validation of IA mock catalogs for DC2 — Production of DC2 mock galaxy catalogs with intrinsic alignments, populated using different empirical models.

📄 DC2 Project: ProtoDC2 Catalog Production — Production steps for protoDC2, a small extra-galactic catalog, to be used for testing pipelines and developing code in advance of the full-sized DC2 catalog.

📄 DC2 Project: Running redMaPPer on DC2 Cosmo and Sim Catalogs — Run redMaPPer cluster finder on DC2 cosmological and image-simulated catalogs,

DESC End-to-End Simulation Workflow

Extra-galactic catalog generation

Image simulations

DM processing

DESC End-to-End Simulation Workflow

○ Responsibility of Cosmological Simulations Working Group   ○ Input   ○ Users

○ Responsibility of Survey Simulation Working Group   ○ Output delivered to collaboration

HACC simulation +
UMachine + HaloTools + Galacticus

PhoSim and ImSim

DM Stack

Extra-galactic catalog generation   Image simulations   DM processing

DC2: Design

- Extragalactic sky: 5000 sq deg input catalog
- Static sky (WL, CL, LSS, PZ) with images: 300 sq deg "main survey" area, 10 years *ugrizy* Wide-Fast-Deep (WFD) cadence

- Time domain (SN, SL) analyses: 1 sq deg "ultra Deep Drilling Field (DDF)" embedded in corner of main survey, 10 years *ugrizy* WFD + DDF visits

- Work led by: H. Awan, R. Biswas, E. Gawiser, P. Marshall et al.

Simulated LSST focal plane   **WFD dithers, Run 2.0**   **DDF dithers, Run 2.0**

DC2: Input Extragalactic Catalog

- HACC "Outer Rim" N-body simulation (more than one trillion particles, 3Gpc/h) provides dark matter distribution (Habib et al., Heitmann et al. in prep.)
- Empirical approach used to populate halos with galaxies to get correct clustering statistics and colors; galaxies then matched with galaxies from semi-analytic model to enrich their properties
- Range of observational data sets is used to tune the model to satisfy diverse DESC science needs at various mass/length scales
- Close collaboration between analysis and computing groups

A. Hearin, D. Korytov, E. Kovacs, P. Larsen, J. Hollowed, S. Rangel et al.

DC2: Simulated LSST Images

PhoSim (Peterson et al.) on NERSC KNL: 1000 nodes allows us to produce 1 full focal plane visit image every 30 seconds - as fast as LSST

11 visits co-added in (r,i,g)

From D. Boutigny

Second image simulation tool; ImSim uses GalSim to produce LSST-specific simulations with tunable levels of complexity.

We use both codes, with a plausible variety of astrophysical and observational effects, to stress-test our analyses

DC2: LSST catalog data, from DM "data release processing" of the images

- Image processing, object detection and measurement, catalog generation will be done by the LSST Facility

- DESC needs its own mirror of the "data release processing" (DRP) pipeline, to reprocess 10% of the LSST images to probe for systematics

- DM-DC2 Task Force is assembling our prototype DRP pipeline, to run at CC-IN2P3 and NERSC

https://docushare.lsst.org/docushare/dsweb/Get/LDM-151

Data Challenge 2: Validation Tools

- Various validation tools are in place to evaluate DC2 data products
  - DESCQA: provides web-based interface to carry out a tests on catalogs at all stages in the DC workflow (Mao et al. ApJS 2018)
  - Exposure checker (adapted from DES; Drliac-Wagner, Krughoff ++)
  - Project's QA tools for DM products
- Tests range from simple "readiness tests" to more elaborate measurements of correlation functions, color-color tests etc.
- Close collaboration with analysis working group to implement tests and derive validation criteria: When is good, good enough for DC2?

2018-04-07_27 :: readiness_protoDC2

Tests completed: instance_test   VALIDATION_TEST_FAILED

Summary outputs:
config.yaml

instance_test   VALIDATION_TEST_FAILED

Outputs:
SUMMARY.html
p00_deg.png
p01_redshift.png
p02_ellipticity.png
p04_size.png
p05_size_bulge.png
p06_shear.png
p07_position_angle.png
p08_convergence.png
p10_sersic_disk.png
p11_sersic_bulge.png

instance_test

redshift_true	0.01205	0.9898	0.8346	0.7211	0
ellipticity_1_true	−0.7749	0.7666	−6.055e−05	−0.0002169	0
ellipticity_2_true	−0.7655	0.8078	2.949e−05	0.0004073	0
size_bulge_minor_true [log]	−4.082	1.041	−0.851	−0.8414	0
size_bulge_true [log]	−3.756	1.195	−0.7203	−0.6999	0
size_disk_minor_true [log]	−4.162	0.681	−1.044	−1.07	0
size_disk_true [log]	−4.015	1.035	−0.8158	−0.836	0
size_minor_true [log]	−3.583	1.5	−0.6392	−0.6172	0
size_true [log]	−3.313	1.69	−0.5195	−0.4793	0
size_bulge_minor_true	8.273e−05	10.99	0.1409	0.2196	0
size_bulge_true	0.0001753	15.68	0.1904	0.2963	0
shear_1	−0.02131	0.0235	0	0.0002582	0
shear_2	−0.0305	0.02803	−0.000375	−0.001083	0
position_angle_true	−90	90	0.09156	0.01601	5.
convergence	−0.02534	0.06704	−0.003825	−0.003941	0
sersic_disk	1	1	1	1	0
sersic_bulge	4	4	4	4	

Data Challenge 2: Extragalactic Catalog Validation Test

Two-point correlation function

ProtoDC2 v3 i-band vs redshift

Data Challenge 2: Validation Tests

- Close collaboration with the Project via the DESC Sensor Anomaly Working Group (SAWG)
  - Comparison of CCD characterization measurements carried out at Brookhaven and SLAC with image simulation tools
  - Excellent opportunity for testing mitigation schemes for sensor defects in DM stack

Flux variation in CCDs due tree rings

PhoSim Flat

Simulation

HY Park, A. Nomerotski and
D. Tsybychev, Properties of
tree rings in LSST sensors,
2017 JINST 12 C05015
Real Data

Data Challenge 2: Validation Tests

Work by Francois Lanusse (CMU)

Simulation after DM

Colors: Different subfields of survey

HSC Y1

Comparison to HSC Y1 shape sample (Mandelbaum et al. 2017)

- Resolution: Indicator of galaxy size compared to size of PSF: R = 1-PSF_size/gal_size
- Galaxy properties look very reasonable!

Data Challenge 2: Validation Tests

Work by Francois Lanusse (CMU)

Simulation (after DM)

HSC Y1

XMM

- **PSF size distribution closest to the HSC XMM field, resolution for that field is quite close to Run1.1p**

- **Effect of blending (and deblending failures) visible in size distribution**

XMM

Example of galaxy with R=0.99 and blendedness = 0.97

Removed by blendedness cut

Simulation!

Resolution distribution

DC2 status

- We are on track to make an *initial internal release* of processed images and DM catalogs for 25 sq deg, 3 season test Run 1.2 by the end of June, along with the full 5000 sq deg "CosmoDC2" extragalactic catalog

- Run 2 (300 sq deg) production will start this summer

- The DC2 data will be used to drive early development of the DESC end-to-end analyses, inferring cosmological parameters from LSST data

Data Challenge 2 Summary

- DC2 provides excellent testbed and development environment for
  - Pipeline creation
  - Systematics investigation
  - Data validation
  - Data access strategy
  - DM stack tests
  - Sensor anomalies
- DM stack generates a lot of data!
  - Data reduction tests (which data products can quickly be re-generated?)
  - DM tools and cookbooks for data access

DC2 provides a complex challenge where the different working groups have to work closely together to establish a science community capable of carrying out a complex joint cosmological analysis on a scale just 10x smaller than LSST

What keeps us up at night …

What keeps us up at night …

- DM stack processing

What keeps us up at night …

- DM stack processing
- Data access and serving

What keeps us up at night …

- DM stack processing
- Data access and serving
- Data validation

DESC welcomes new members!