The Dark Energy Survey Project Office

DARK ENERGY SURVEY

The Dark Energy Survey

The Cycle 5 Challenge Plan: Simulation, Data, and Analysis

Version 1 DRAFT

James Annis
Huan Lin
Joe Mohr
Josh Frieman
Ofer Lahav

22 May 2009

CONTENTS

1 INTRODUCTION

The challenge cycle of simulations, data reduction, and science analysis and testing is central to our preparations for the science of the DES.

This document is in part to describe the simulation, data reduction, and science analysis challenge cycle and in part to write down specific data quality goals. The process of preparing the document included science working group input and the science requirements document, the simulation development planning, and the goals of DESDM in the data challenge. Specifying specific goals will also rationalizes the testing work of the science working groups during the analysis challenge.

The document also serves to document our requirements for access to the science data and science codes by the collaboration during the 5th analysis challenge.

2 SCIENCE OBJECTIVES

The three overall goals of this challenge cycle are to improve the measurements of shape, the measurements of photometry, and the measurement of the survey mask.

2.1 SCIENCE GOALS

2.1.1 The Main Survey

The interest in improving shape measurement quality is driven by the desire to perform an astrophysically meaningful shear signal measurement using the simulated data. The cluster and weak lensing working groups desire, for example, to measure the weak lensing signal from the ensemble of clusters found. The interest in improving photometric quality is driven by a desire to improve the photometric redshifts, of great import to weak lensing, clusters, and large scale structure. The interest in improving the knowledge of the survey mask is driven by the desire of the large scale structure working group to perform a full scale power spectrum analysis using using masks derived from the data set.

2.1.2 The Supernova Survey

The supernova working group wishes to measure its SN detection efficiency as a function of magnitude inside the simulated data set.

1. Improve the SN image simulations to include backgrounds including asteroids and non-Ia SN transients.
2. Generate and process a full season of SN data for one SN field (this is the same amount of data as we are doing for DC-4).
3. Achieve a SN detection efficiency of >50% and measure the asteroid rejection efficiency.
4. Serve the results of the SN object finding via the Oracle Database & Apache web server using SN-specific web pages, the generic sql interface, and custom C-language interface programs.
5. Test SN-specific monitoring and quality control tools developed by the SN group.

The emphasis is on real-time data processing for SN identification for followup observing.

2.2 SCIENCE REQUIREMENTS

The whole of the SRD and the DESDM TRD provide requirements that will need to be met by the time of science commissioning. Any and all of these should be tested at the end of the Cycle 5 process.

For the purposes of Cycle 5, however, we focus on the two of the overall goals of the challenge cycle, improving the measurements of shape and photometry.

In image quality, we focus on three requirements.

1) The median PSF FWHM (averaged over all exposures and over the survey area) in each of the r, i, and z bands should be less than 0.9”
2) The mean PSF whisker length for stars/exposure must be below 0.2” in the r,i,z bands for the wide area survey.
3) For measured shear $\gamma_i^{meas} = (1+m)\gamma_i^{true} + \gamma_i^{add}$. We have a requirement that m<0.004 and $\gamma_i^{add} < 4\times10^{-4}$. For the purposes of this cycle, we set as a goal $\gamma_i^{add} < 1\times10^{-3}$.

In photometric measurements quality, we focus on six requirements:

1) Limiting magnitudes of grizY= 24.0, 24.0, 24.0, 23.6, 21.6
2) 2% photometric calibration
3) Types of magnitudes: aperture magnitudes, PSF magnitudes, total magnitudes (e.g. SDSS Petrosian mags), color optimized magnitudes (e.g. SDSS cModel mags)
4) Star-galaxy separation must use the position dependent model of the PSF and be accurate at >= 95% at the 10 sigma photometric limit.
5) Aperture corrections are necessary for calculating magnitudes.

3 SIMULATION CHALLENGE GOALS

The simulated area will again consist of 100-200 sq-degrees of sky to the full depth of the survey.

An ancillary simulated dataset is one tile observed to the full depth of the survey, organized as if it was taken during one night, along with enough data to produce the overlapping adjacent tiles. This is called the "golden night". It consists of roughly 52 science images plus calibrations and flat fields.

3.1 CATALOG LEVEL GOALS

The Simulation working group will produce new mock galaxy catalogs with increased realism in the physics of galaxy evolution. These catalogs will have shapes associated with the galaxies. The shapes will be sheared by a model of the intervening mass. Required is shear from the intervening dark matter halos in the halo catalog produced by the simulation group, assuming shear from a SIS or NFW profile. A goal is to have the shear from the underlying cosmic density field in the dark matter simulation.

The Milky Way study group will produce a star catalog for use in the simulations that incorporates the same DES system response curves used to produce the galaxy colors. The brighter stars ($i < 20$) will be consistent with the USNO-B catalog, and the fainter stars from a physically meaningful galactic structure model.

The Survey Strategy team will produce an observing plan for the data that mimics baseline survey strategy.

3.2 IMAGE LEVEL GOALS

3.2.1 Photometry

The DECam system will have a spatially varying system response due to the effects of the quantum efficiency curve differing from CCD to CCD and due radial variations in the filters. Spatial variation of the system response will be modeled as a color dependent correction to the fluxes of objects.

The sky transparency is spatially dependent under slightly non-photometric conditions. This will be modeled as a quadratic polynomial in focal plane x and y.

The sky brightness is spatially dependent for several reasons, the largest of which is the presence of the moon. Sky brightness will be modeled as a quadratic polynomial in focal plane x and y.

Flat fields differ from chip to chip. This will be modeled using a variety of flat fields gathered from CCD testing lab. Calibration flat fields will be provided that differ from the underlying simulation flat fields.

The DECam electronics exhibit crosstalk between amplifiers on the same CCD at the 0.2% level and between CCDs on the same electronics board at the 0.1% level. Cross talk between CCDs sharing the same card will be modeled.

3.2.2 Shape

Pixel size variations will be modeled. The model for the geometrical distortion due to the optics will be upgraded to include the size change in the pixels. Likewise the CCD itself has variable pixel sizes due to the electrostatic field inside the chip bending at the array edges ("the glowing edges") and column to column periodic size changes and these will be modeled. The effects of distortion due to differential refraction by the atmosphere will be included (note that this primarily affects astrometry).

The PSF library will upgraded to include PSFs appropriate to small amounts of defocus and misalignment.

Galaxy shapes realized from a shapelet catalog derived from high resolution HST data will be used. The 2 sq-degree HST COSMOS data will allow more better galaxy shape coefficients to be determined and thus the galaxy shapes in the simulations will be more reliable.

The eight focus and alignment chips will have their output simulated.

4 DATA CHALLENGE GOALS

4.1 ASTRONOMY CODES

4.1.1 Photometry

New model fitting photometry capabilities have been developed within SExtractor during DC4. For DC5 we expect to employ the model fitting photometry for all the coadd images. We will explore whether having model fitting photometry on the single epoch images is required to achieve any of our data quality requirements.

We will integrate and test new star/galaxy classification capabilities available within SExtractor and apply them to the DC5 dataset.

We will develop and test an algorithm to estimate a source subtracted background within SExtractor and employ that in the DC5 cataloging. We expect this algorithm to resolve the slight background overestimation that we have seen in the DC4 testing.

PSF homogenization code was developed and integrated into a DC4 prototype coaddition pipeline. Significant work remains to tune this code and understand its effects on completeness and photometric measurement uncertainties. We will continue our efforts to develop this functionality for the project, because it goes hand in hand with using the new SExtractor model fitting photometry on coadds.

Effort will be invested to improve the image by image scaling of the fringe correction.

We are developing code to create a pupil ghost from calibration data.

We are developing code to extract the crosstalk coefficients directly from the DECam observations.

We will develop a code to produce star flats from the DECam photometric catalogs and incorporate a star flat correction into the core image processing.

4.1.2 Shapes

We are working to improve image photometric flatness within the DESDM pipelines by developing a new flatness correction that accounts for the differences in pixel scale over the CCD during the flat fielding process.

4.1.3 Masking

We will continue our work on image masking, extending the existing codes to include masking of satellite trails. We expect to employ the Hough transform, which is well suited to detecting linear features in data. We will continue to improve our stellar masking code that removes diffraction spikes and bleed trails as well as the bright halo surrounding the brightest sources. Additional testing and tuning of the DECam cosmic ray masking using the Terapix eye and SExtractor software will be carried out.

We will adopt a new mode of calculating weight maps that do not include the noise contributions from object flux. These new weight maps are expected to provide improved quality coadd images and to be more useful for generating survey depth as a function of the position,

4.2 SCIENCE CODES

4.2.1 Weak Lensing Pipeline

In DC4 we developed and used the first DES weak lensing pipeline. But this is just a first step, because it calculated shear from each single epoch image independently. For DC5 we expect to test a version of the pipeline that carries out joint shear measurements using all images available (selected by band, seeing, depth) for a given object. We will continue to work with DM member M. Jarvis and other members of the Weak Lensing SWG (Jain, Bridle, Sheldon et al) on this project. Our goal is to run a pipeline in DC5 that produces scientifically useful data for the WL SWG.

4.2.2 Difference Imaging Pipeline

We will continue to work with the SNe SWG (Marriner, Kessler, Fisher et al) in the development of a difference imaging pipeline. The plan is for this pipeline to be used by the DESDM operations team to carry out difference imaging within the broader DES survey, and by the SNe SWG to carry out the difference imaging within the SNe science fields. Our goal is to run a pipeline in DC5 that produces scientifically useful data for the SNe SWG.

4.2.3 Large Scale Structure Analysis Masks

We are working with the LSS SWG (Percival, Gaztanaga, Swanson et al) to create a pipeline that will calculate survey masks in a form that is required for the analysis of clustering data. There is significant design work remaining, but the idea is to work from either DC5 image metadata or the weight maps for the DC5 coadds to build the data products needed to understand the variations in depth as a function of position within the survey. This is tracked at the pixel level in the coadd images, but a higher level and more compressed data product is needed to enable efficient analysis of the DC5 galaxy clustering.

5 ANALYSIS CHALLENGE TECHNICAL REQUIREMENTS

In order for the science collaboration to analyze the data challenge data set, or indeed the real DES data set, there are a series of technical requirements that should be met.

5.1 ACCESS TO THE PORTAL

SWG members will have access to a portal on the DESDM database which can perform efficient queries of large numbers of selected quantities from calibrated object tables and return the output (which may be up to several Gigabytes per query) in ASCII (CSV, TSV) format to the SWG member's home machine for further analysis.

Efficient means: indexed on at least RA,DEC but also on such things as run number, color, magnitude in several magnitude systems. It also implies that if necessary, 'stored procedures' would be developed to pre-calculate 'science object tables', derived from the join of the astronomy code tables (SExtractor catalogs) and the science working group code tables (i.e. WL, photo-z), if necessary, which would be updated frequently, perhaps whenever a co-add table was generated.

Timescales for queries to return typical volumes should be specified. Some example queries or query patterns, specified by the SWG or Project Scientist, would help.

Evaluation of the portal in the light of precursor work such as SDSS-CAS, CASJOBS, and MYDB should be performed.

5.2 AUTHORIZATION

An Authorization mechanism to the database via a portal from the outside should be specified and allow for 'power user access' during the data challenges. i.e. can Authorization be 'command line scriptable', so that one doesn't have to enter a user/passwd by hand into a browser window for each query of the database.

5.3 INTERFACE/DESIGN SPECIFICATIONS

Interface and/or design specifications should exist:

1) for data
    a. input and output file formats
    b. grouping requirements [i.e. can you run on one CCD frame in one filter, or do you need a whole set of 62 CCDs, or do you need all to-be-coadded observations over the same part of the sky, plus these auxillary catalogs]
2) for code
    a. must it compile on what Linux node with what gcc libraries, 32/64 bit

issues, multi-thread capability issues or restrictions

3) for parameter/control files
    a. files which control tunable parameters in the code
    b. control where the code can find its data and where it can run, etc.

5.4 UNIT TESTING

Individual sections of science code, including astronomy codes (Sextractor) and SWG codes (WL, difference pipeline) should be 'unit testable' using small subsets of the data (simulated or real), which may be extracted from the large DESDM database, or provided by the code developers.

5.5 CODE REPOSITORY AND VERSIONING

All pipeline code (astronomy codes and SWG codes) should be checked in to some sort of version control system (i.e. SVN) with increasing version numbers. All collaborators should eventually have read access to all pipeline source code.

When a data challenge test is run which crosses ‘institutions’ (i.e. involves more than one working group or the DESDM and a working group), a 'tagged' version of code should exported from the appropriate repository is used. There may be more than one repository at more than one institution, at least initially.

5.6 CODE CHANGE CONTROL

All DES software code have some change control mechanisms, aiming for a balance between avoiding locking down versions of code too early before a data challenge and avoiding changes at the last minute which break other connected pieces of a pipeline during a mass-processing session.