The LSST DESC Data Challenge 1: Simulating data for the next generation of photometric redshift surveys

Yusra AlSayyad,[1] Humna Awan,[1] Colin Burke,[1] Jun Cheng,[1] James C. Chiang,[1] Scott Daniel,[1] Seth Digel,[2] Richard Dubois,[1] Eric Gawiser,[1] Tom Glanzman,[1] Mike Jarvis,[1] Tony Johnson,[1] Heather Kelly,[1] David Kirkby,[3] Simon Krughoff,[1] Robert Lupton,[1] Rachel Mandelbaum,[1] P. Marshall,[4] Mustafa Mustafa,[1] En-Hsin Peng,[1] John Peterson,[1] Paul Price,[1] Javier Sanchez,[5] Glenn Sembroski,[1] Anze Slosar,[1] Brian Van Klaveren,[1] Chris W. Walter,[1] Matt Wiesener,[1] and Bo Xin[1]

(LSST Dark Energy Science Collaboration)

[1] *Affiliation*
[2] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[3] *Deparment of Physics and Astronomy, University of California, Irvine, CA, USA*
[4] *Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[5] *Department of Physics and Astronomy, University of California, Irvine, CA, USA*

The success of future Stage IV dark energy surveys (Albrecht et al. 2006) relies in the ability to model and mitigate systematic uncertainties. Realistic simulation offer a unique opportunity to study systematic uncertainties and test the processing and analysis pipelines of ongoing and future experiments. Here we present a set of realistic simulations of ~ 40 sq.-deg. that try to mimic the depth and characteristics of LSST 10-years coadd images in the r-band. We characterize our samples performing several astrometric and photometric checks to assess the quality of the measurements and to enable the usage of these simulations for future studies.

Introduction

The increase in statistical power from recent cosmological experiments makes the modeling, and mitigation of systematic uncertainties key to extract the maximum performance and produce competitive analyses. More traditional in high energy

particle physics (Brun et al. 1978), (Sjöstrand et al. 2006), end-to-end simulations provide a unique framework to model systematics and streamline processing and analysis pipelines given that we have complete information about the inputs and outputs. With the larger availability of computational resources this approach has also been extended to photometric redshift galaxy surveys (Suchyta et al. 2016; Bruderer et al. 2016) and a similar effort is undergoing in spectroscopic surveys such as DESI (DESI Collaboration et al. 2016). For surveys like the LSST (Ivezic et al. 2008) where the expected data volume is very large and where a highly stringent control of the systematic uncertainties is required, producing these kind of end-to-end simulations becomes necessary.

In this paper, we present the procedure to generate and process images that resemble the data that will be produced by LSST (Ivezic et al. 2008) after 10 years of operation in r-band using state of the art tools. We also characterize the products of this process for future studies. These productsencompass single-visit and coadded calibrated exposures (i.e., flattened, background removed, etc) and source catalogs that add up to $\sim 225TB$. They are the result of three different simulations: imSim dithered, imSim undithered, and PhoSim that will be introduced later.

This paper is structured as follows: In Section Section 2 we describe the input catalog used for our simulations, in Section Section 3 we introduce two different approaches to generate simulated images to resemble LSST data. In Section Section 4 we present the procedure and tools used to perform calibration and source extraction on the simulated images. In Section Section 5 we describe the output catalogs produced by our pipelines. Finally, in Section Section 6 we present some concluding remarks.

Image generation: inputs

Describe CatSim inputs and dithering

Image generation: pipeline

The artificial generation of astronomical images is a very complex and computationally demanding process. In the recent years there is a big effort in the community in order to create software that allows more realistic and fast image generation (Suchyta et al. 2016; Bruderer et al. 2016). In our case, we use two different approaches: In one approach we use modeling of the input sources using GalSim (Rowe et al. 2015). The other approach consists in running a full photon-

shooting simulation using PHOSIM (Peterson et al. 2015). The former has a big speed advantage but the latter fully traces each photon coming from the sources through the atmosphere and the instrument, increasing the level of realism. These two approaches allow us to focus on different systematic effects and science cases.

The imSim pipeline

Describe imSim

The PhoSim pipeline

Describe PhoSim

Image processing pipeline

Once the images are produced we process them using the LSST software stack (Jurić et al. 2015). This is an open source high-performance data processing and analysis system intended for use in O/IR survey data. The code can be found at `dm.lsst.org` and `pipelines.lsst.io`. The raw, uncalibrated single exposures are used as inputs. The software performs the reduction, detection, deblending and measurement on individual visits and coadds producing the level 2 data products (Jurić et al. 2015).

Say something about data size, times, configuration, etc

Output catalogs

After being processed, the catalogs are accessible by DESC collaborators and stored at NERSC. We generate pandas dataframes and three different databases for each one of the total coadd catalogs in order to be accessed by the collaborators and perform their own analyses. These catalogs contain 10.6 million objects covering an area of $\sim$ 43 deg^2.

In order to check the level of realism and the accuracy of the processed catalogs we perform several quality assurance tests. We focus on three different areas that

can induce a systematic effect in the weak lensing and clustering observables: astrometry, photometry and PSF.

Astrometry checks

Biases on astrometry can potentially affect both clustering and weak lensing measurements. These biases can have different origins: PSF mis-characterization, not corrected sensor effects, presence of blended sources are among the most common scenarios for single-visit exposures. In the case of co-adds we should add to this list a different effect: incorrect modeling of proper motion for the measured objects.

We will follow two approaches to check the quality of the astrometric solutions that we obtained: an *external* check comparing to the input *truth* catalog; and an *internal* check comparing different visits.

External checks

As we have already mentioned, one of the big advantages of using simulations is that we have access to the *true* underlying information. We will use this information to check the precision of the astrometric measurements in single exposures and co-adds. In both cases we will use a `KDTree` (Pedregosa et al. 2011) to retrieve those objects in the input catalog that are in a radius of 0.2 arc-seconds of those detected in the output catalog and select the match that is closest in magnitude. We only consider sources which have a magnitude difference smaller than 0.05 magnitudes.

We selected a representative single visit and calculated the difference between the measured and the input positions. These are represented in Figure Figure 1. We can see that both RA and Dec distributions are compatible with each other, meaning that there are no anisotropies in the detection, as expected from the inputs. However, we find that the distributions are assymetric and that the median is not zero. This effect is even more noticeable when we accumulate all of the visits and it is present in the dithered and undithered runs.

Internal checks

For the internal checks we select two
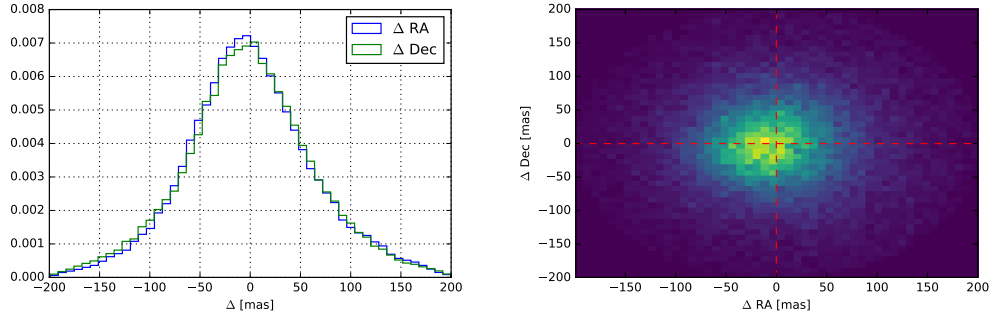
Figure 1. Distribution of the difference $\Delta = X_{measured} - X_{input}$ in RA (blue) and Dec (green) coordinates. We cannot appreciate any differences between these, however we see that there is median is not at zero $\Delta_{median} \approx -2$ mas. The histograms are normalized such that the total sum of the counts is equal to one. We selected one random representative exposure (visit number 270675 for the IMSIM dithered run).

Photometry checks

PSF checks

Conclusions

Acknowledgments

Here is where you should add your specific acknowledgments, remembering that some standard thanks will be added via the `acknowledgments.tex` and `contributions.tex` files.

This is the text imported from `acknowledgments.tex`, and will be replaced by some standard LSST DESC boilerplate at some point.

Author contributions are listed below.
Yusra AlSayyad: Contribution
Humna Awan: Contribution
Colin Burke: Contribution
Jun Cheng: Contribution
James C. Chiang: Contribution
Scott Daniel: Contribution
Seth Digel: Contribution
Richard Dubois: Contribution
Eric Gawiser: Contribution
Tom Glanzman: Contribution

Mike Jarvis: Contribution
Tony Johnson: Contribution
Heather Kelly: Contribution
David Kirkby: Contribution
Simon Krughoff: Contribution
Robert Lupton: Contribution
Rachel Mandelbaum: Contribution
P. Marshall: Contribution
Mustafa Mustafa: Contribution
En-Hsin Peng: Contribution
John Peterson: Contribution
Paul Price: Contribution
Javier Sanchez: Validation and analysis, started note
Glenn Sembroski: Contribution
Anze Slosar: Contribution
Brian Van Klaveren: Contribution
Chris W. Walter: Contribution
Matt Wiesener: Contribution
Bo Xin: Contribution

References

Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, ArXiv Astrophysics e-prints, astro-ph/0609591

Bruderer, C., Chang, C., Refregier, A., et al. 2016, ApJ, 817, 25

Brun, R., Hagelberg, R., Hansroul, M., & Lassalle, J. C. 1978, Simulation program for particle physics experiments, GEANT: user guide and reference manual (Geneva: CERN)

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, ArXiv e-prints, arXiv:1611.00036

Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, ArXiv e-prints, arXiv:0805.2366

Jurić, M., Kantor, J., Lim, K., et al. 2015, ArXiv e-prints, arXiv:1512.07914

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Peterson, J. R., Jernigan, J. G., Kahn, S. M., et al. 2015, ApJS, 218, 14

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121

Sjöstrand, T., Mrenna, S., & Skands, P. 2006, Journal of High Energy Physics, 5, 026

Suchyta, E., Huff, E. M., Aleksić, J., et al. 2016, MNRAS, 457, 786