

The LSST DESC Data Challenge 1: Generation and Analysis of Synthetic Images for Next Generation Surveys

J. SÁNCHEZ,^{1,2} C. W. WALTER,³ A. SLOSAR,⁴ D. KIRKBY,⁵ J. CHIANG,⁶ T. GLANZMAN,⁶ S. F. DANIEL,⁷ H. AWAN,⁸ E. GAWISER,⁸ W. M. WOOD-VASEY,⁹ Y. ALSAYYAD,¹⁰ C. BURKE,^{11,12} J. CHENG,¹¹ S. DIGEL,⁶ R. DUBOIS,⁶ M. JARVIS,¹³ T. JOHNSON,⁶ H. KELLY,⁶ S. KRUGHOFF,¹⁴ R. H. LUPTON,¹⁰ R. MANDELBAUM,¹⁵ P. J. MARSHALL,⁶ M. MUSTAFA,¹⁶ E. -H. PENG,¹¹ J. R. PETERSON,¹¹ P. PRICE,¹⁰ G. SEMBROSKI,¹¹ B. VAN KLAVEREN,⁶ M. P. WIESNER,¹⁷ AND B. XIN¹⁸
(LSST DARK ENERGY SCIENCE COLLABORATION)

¹*Department of Physics and Astronomy, University of California, Irvine, Frederick Reines Hall, Irvine, CA, U.S.A.*

²*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL, U.S.A.*

³*Duke University, Department of Physics, Durham, NC, U.S.A.*

⁴*Brookhaven National Laboratory, Upton, NY, U.S.A.*

⁵*Department of Physics and Astronomy, University of California, Irvine, Frederick Reines Hall, Irvine, CA*

⁶*SLAC National Accelerator Laboratory, 2575 Sand Hill Rd, Menlo Park, CA, U.S.A.*

⁷*Department of Astronomy, University of Washington, Seattle, WA, U.S.A.*

⁸*Department of Physics & Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Rd, Piscataway, NJ U.S.A.*

⁹*Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC). Physics and Astronomy Department, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.*

¹⁰*Princeton University, Princeton, NJ, U.S.A.*

¹¹*Department of Physics and Astronomy, Purdue University, West Lafayette, IN, U.S.A.*

¹²*Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL, U.S.A.*

¹³*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, U.S.A.*

¹⁴*LSST Project Management Office, Tucson, AZ, U.S.A.*

¹⁵*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA, U.S.A.*

¹⁶*Lawrence Berkeley National Laboratory, One Cyclotron Rd, Berkeley, CA U.S.A.*

¹⁷*Benedictine University, 5700 College Road, Lisle, IL, U.S.A.*

¹⁸*Large Synoptic Survey Telescope, Tucson, AZ, U.S.A.*

ABSTRACT

Data Challenge 1 (DC1) is the first of the planned three synthetic datasets produced by the Large Synoptic Survey Telescope (LSST) Dark Energy Science Collaboration (DESC). These are designed to develop and validate data reduction methodologies and to study the impact of systematic effects that will affect the LSST dataset. DC1 is comprised of r -band observations of 40 square degrees to 10 year LSST depth. We present how the dataset was: a) generated, by simulating sources from cosmological simulation in individual sensor-visit images with different observing conditions; b) reduced using a development version of LSST Science Pipelines; and c) matched to the input catalog. We study performance of our processed catalogs compared to the LSST requirements on key performance metrics (KPMs). We establish a set of pipeline flags that produce a clean extragalactic sample and discuss residual sample contamination, including efficiency of star-galaxy separation and deblending algorithm. We demonstrate the power spectrum measurement on the simulated field. Our main conclusions are: i) realistic and validated synthetic datasets will be required for a successful control of systematics in the upcoming photometric dark energy experiments; ii) within the fidelity of DC1 image generation, the LSST Science pipelines passes all testable KPMs in DC1; iii) there is no individual noiseless method for matching the input and output catalogs; iv) lack of multi-band observations significantly hampers sample definition and cleaning in DC1.

Keywords: large-scale structure of the universe

1. INTRODUCTION

The increase in statistical power from recent cosmological experiments makes the modeling and mitigation of systematic uncertainties key to extracting the maximum amount of information from these surveys. 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 since we have a complete understanding of the inputs and outputs. With the increasing availability of computational resources, this approach has also been extended to imaging surveys (Bruderer et al. 2016), and similar efforts are being undertaken in spectroscopic surveys such as DESI (DESI Collaboration et al. 2016).

For surveys like the Large Synoptic Survey Telescope (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 enables successful validation and verification of the processing and analysis pipelines. With an expected total of ~ 50 PB of raw data and ~ 40 billion objects (Ivezic et al. 2008) after 10 years, the data handling becomes challenging even when considering a small subset of the LSST area.

The LSST Dark Energy Science Collaboration (DESC¹) has planned a series of Data Challenges (DCs) carried out over a period of years, aimed at successively more stringent and comprehensive tests of analysis pipelines, to ensure adequate control of systematic uncertainties for analysis of the LSST data. An additional goal of these DCs is development of the infrastructure for analyzing, storing, and serving substantial data volumes; even while using the outputs of the LSST Science Pipelines as inputs into analysis pipelines, the analysis pipelines will need to handle quantities of data beyond those seen by ongoing surveys, even after just a single year of LSST data is available. Moreover, it is anticipated that non-negligible subsets of the data may need to be reprocessed to generate systematic error budgets (e.g., assessing sensitivity of the results to certain stages of the analysis process by changing some parameters in the analysis). These goals will be achieved in practice by a combination of reprocessing of the precursor datasets, which have the advantage of being realistic; and completely synthetic datasets, that have both a known ground truth and the ability to turn various ef-

fects on and off and thus study them in a more controlled environment.

Both goals of the DCs dictate a gradual increase in the sophistication and volume of the simulated data. In this paper, we present and analyze simulated images from DC1, the first of three such data challenges planned within DESC. The nominal goal of this dataset is to produce synthetic data corresponding to 10 years of integration in the r -band over a contiguous patch of the sky covering approximately 40 deg^2 . This represents just one of the LSST's six filters and covers a rather small fraction ($\sim 0.2\%$) of the total LSST area. We describe how this was achieved and characterize the resulting products. We validate the basic photometric and astrometric calibration of these products and check the performance of the pipeline against the requirements set by LSST and DESC in their respective Science Requirements Documents (Ivezić & the LSST Science Collaboration 2013; The LSST Dark Energy Science Collaboration et al. 2018). These data products encompass single-visit and coadded calibrated exposures (i.e., flattened, background subtracted, etc.) and source catalogs that, together, take up to ~ 225 TB of disk space. To check suitability of this dataset for galaxy clustering measurements, we perform a two-point clustering analysis in harmonic space, assess the impact of different observing conditions/foregrounds as potential sources of systematic effects and how dithering can mitigate this impact.

This paper is structured as follows: Section 2 includes a summary of the factors that informed the design of this data challenge. In Section 3, we describe the inputs for our simulated images. In Section 4, we discuss the dither strategies used for this study. In Section 5, we describe the process used to generate, process and reduce LSST-like artificial images. In Section 6, we describe the processed data products generated and perform several validation tests. In Section 7 we describe the procedure to obtain a clean data sample suitable for clustering analyses. In Section 8, we present the clustering analyses on the simulated data products. Finally, in Section 9, we present some concluding remarks.

2. DATA CHALLENGE DESIGN

As mentioned in Section 1, the design of this first data challenge (DC1) is driven by a combination of several factors: the need to test analysis pipelines for control of systematic uncertainty, and to develop infrastructure for processing and serving data in a way that is useful to DESC. The philosophy behind the design of the DESC data challenges is to increase the complexity and level of realism of the datasets in each subsequent iteration.

¹ <http://lsstdesc.org/>

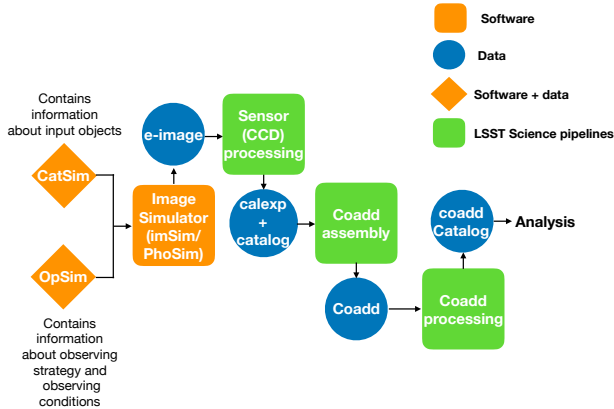


Figure 1. Workflow diagram for full data simulations. The image simulation step produces synthetic raw observations from a known truth catalog based on N-body simulations and simulated observing conditions. These data, in the same format as actual observations from the LSST telescope are processed by the data management to first generate calibrated single exposure images. These are calibrated both astrometrically and photometrically and are fed again into data management stacks that produces the image co-adds and a catalog of detected static sources.

Thus, DC1 is limited in scope and focus, testing a subset of the systematics affecting the different probes that will be used in DESC.

As a stepping stone to eventually producing images covering hundreds to thousands of square degrees in DC2 and DC3, DC1 covers a 40 deg^2 footprint. This is sufficient to enable tests of two-point clustering statistics up to ~ 1 degree scales. To ensure the simulated image volume is tractable, DC1 only includes images in a single band (r -band), but goes to full LSST 10-year depth.

The full simulation workflow is depicted in Figure 1. Briefly, we use as inputs the position, shapes and fluxes from a galaxy mock catalog from **CatSim** (Connolly et al. 2010, 2014), which we describe in more detail in Section 3, and the observing conditions and strategies described in Section 4 using **OpSim** (Delgado et al. 2014). These are passed to our image simulation packages describe in Section 5.1 that produce raw e-images (i.e., full sensor images without any added instrumental effects such as cross-talk, bleeding, etc.). These e-images are then processed by the LSST Science Pipelines (Jurić et al. 2015; Bosch et al. 2018). The processing is described in Section 5.2. After this, we obtain the calibrated exposures, coadds and catalogs that we use for our analysis.

Image simulations allow us to study in detail the detection and deblending properties of a given image-processing pipeline. For example, if we produce images using an object catalog with random positions uniformly distributed across the sky, as well as uniformly random shapes and fluxes, we can get information about detection efficiencies as a function of flux. However, the information about blending will not be realistic and we will not be able to capture some correlations present in real data. On the other hand, using N-body simulations as the input to generate artificial images allows us to study all the aforementioned effects. This is why we used the **CatSim** (Connolly et al. 2010, 2014) catalog as our input. **CatSim** is a set of simulations provided by the LSST Simulations Team representing a realistic distribution of both Milky Way and extra-galactic sources. In particular, the extra-galactic catalog contains galaxies covering the redshift range $0 < z < 6$ in a 4.5×4.5 degree footprint. The galaxies are generated by populating the dark matter haloes from the Millennium simulation (Springel et al. 2005) using a semi-analytic baryon model described in De Lucia et al. (2006) including magnitudes BVR*i*K and bulge-to-disk ratios. For all sources, a spectral energy distribution (SED), is fit to the galaxy colors using Bruzual & Charlot (2003) spectral synthesis models. Fits are undertaken independently for the bulge and disk and include inclination dependent reddening. Morphologies are modeled using two Sérsic profiles (Sérsic 1963) and a single point source (for the AGN). Half-light radii for the bulge components are derived from the absolute-magnitude vs half-light radius relation given by Gonzalez et al. (2011). Stars are represented as point sources and are drawn from the Galfast model (Jurić et al. 2008). More information about these catalogs can be found at the LSST Simulations webpage².

For the Data Challenge 1 (DC1), we chose a nominal field centered at $\text{RA} \approx 93^\circ$ and $\text{Dec} \approx -29^\circ$. This field has a galactic latitude of $b \approx -20^\circ$ and a dust extinction $0.05 \leq E(B - V) \leq 0.35$ and thus represent a typical LSST dataset. The **CatSim** catalog was tiled to generate a $\sim 40 \text{ deg}^2$ footprint covering 4 LSST full focal plane pointings. This approach introduces a periodicity that induces extra correlations in our sample, however, this is not a major defect as we are unable to measure correlations on relevant scales ($> 1 \text{ deg.}$) with any useful precision.

After tiling, the input catalog contains approximately 63.1 million sources, of which 97% are galaxies whose

3. IMAGE GENERATION: INPUT CATALOG

² <https://www.lsst.org/scientists/simulations/catsim>

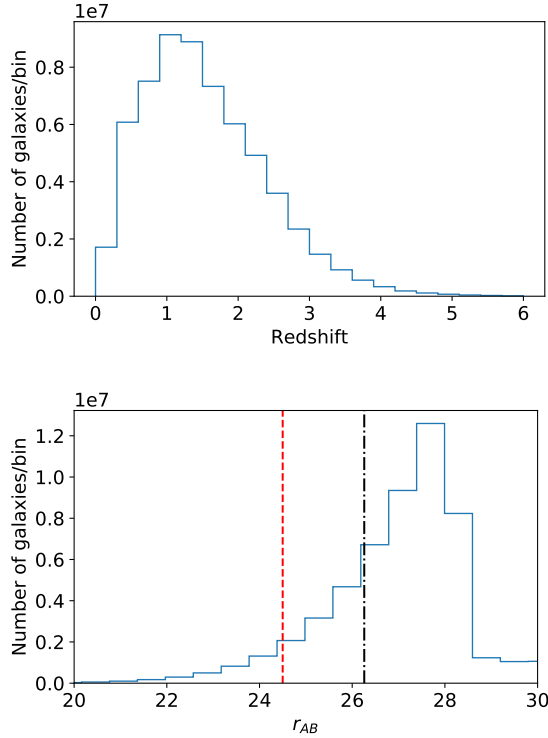


Figure 2. Redshift (top) and magnitude (bottom) distribution for the galaxies used as inputs for the Data Challenge 1 simulations. In the magnitude distribution we include, as references, the typical depth for a single exposure (red dashed line) and the median depth in the deepest coadded DC1 simulation (black dashed-dotted line).

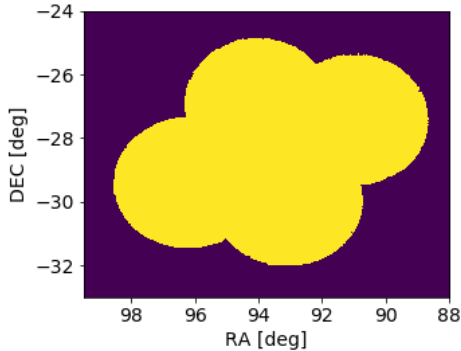


Figure 3. Footprint of the DC1 dataset. We simulate 4 LSST full focal plane pointings which roughly corresponds to 40 deg^2 .

redshift and magnitude distributions are depicted in Figure 2 and remaining objects are stars. We simulate images in r -band to the LSST full depth (10 years). The final footprint can be seen in Figure 3. We simulate observations within this footprint using a simulated observing cadence generated with the LSST Operations

Simulator (OpSim) (Delgado et al. 2014). In particular, we use the `minion_1016`³ database, which contains simulated pointing position, observation date and filter. It also contains information about simulated observing conditions, such as, seeing, sky brightness, moon position, etc. We use ≈ 4 pointings with ≈ 184 visits per pointing over 10 years from this database for DC1.

4. DITHER STRATEGY

As mentioned in Section 2, we use OpSim’s output, which contains a realization of the LSST observing cadence and the survey footprint. Since OpSim divides the sky with hexagonal tiles, the nominal telescope pointings lead to overlapping regions across adjacent tiles that are observed more often than the non-overlapping part of the field-of-view (FOV), resulting in depth non-uniformity on the scale of ~ 1 degree and consequent systematic uncertainties (Awan et al. 2016). In an effort to mitigate these effects, and following the same approach that will be taken with LSST data, we implement *dithers* – offsets in the nominal telescope pointings. Specifically, here we use *large*, i.e., as large as the FOV, random translational dithers, implemented after every visit, and random rotational dithers implemented after every filter change. The specific translational dither strategy is chosen based on a more extensive study of the various (translational) dither strategies in Awan et al. (2016), where random dithers after every visit are found to be amongst the most effective.

For our purposes, we consider both an undithered and a dithered observing strategy. For the dithered strategy, some visits will contain sensors that fall out of the DC1 region; these sensors were not simulated in order to save computational resources. However, the sensors that partially overlapped our nominal field of view were simulated. In total, we simulate ≈ 184600 sensor/visits for the dithered simulation and ≈ 151000 for the undithered simulation. (AS: But for dithered option, presumably the right thing to do would be also considered pointings that are not nominally on our field, but happened to be partially dithered into it. Otherwise, the total depth of the undithered is higher as you never throw anything away? Explain.)(JS: I think this is the confusion between sensor and visit. The visits with sensors within the FOV were simulated. The sensors partially overlapping the footprint were simulated. However, the sensors of these visits that didn’t overlap the footprint, were not simulated (so we could save some CPU hours).)

³ <https://www.lsst.org/scientists/simulations/opsim/opsim-v335-benchmark-surveys>

5. IMAGE GENERATION AND PROCESSING

The artificial generation of astronomical images is a complex and computationally demanding process. In recent years, there have been major efforts in the community to create software that enables the generation of astronomical images, with various choices made in terms of level of complexity and fidelity, and computational efficiency, such as UFIG (Bruderer et al. 2016), and PHOSIM (Peterson et al. 2015). In our case, we model the input sources using two independent approaches. Firstly, we use PHOSIM, a fast photon Monte Carlo code that enables the generation of images with a high-level of realism. We use a custom configuration that enables some approximations on the sky background in order to reduce the overall computing time needed to produce the images.

We also employ imSim⁴ (Walter et al., In prep.), which internally uses GALSIM (Rowe et al. 2015) as a library for image rendering, and uses LSST-specific information (e.g., about the geometry of the CCDs and the focal plane, the system throughputs in the different bands, etc.) to generate LSST-like images. imSim, in the configuration used for DC1, performs a series of approximations that allow us to complete the image simulations significantly faster ($\sim 60\times$) than PHOSIM, at the expense of losing realism. In this work, we will focus on the results obtained from the images generated by imSim.

(AS: Are we completely ignoring phosim runs here then? I think we should at least motivate imsim as a fast version of phosim. I'd say we did both, but say that in this first paper we are focusing on imsim results)

5.1. imSim

imSim is an open-source image simulation software package that uses GalSim with a modular approach. It allows the user to change the level of realism of the simulations by changing the complexity of PSF, background and source modeling. In particular, we used a pre-release version specifically meant to perform the DC1 simulations: imSim v.0.1.0⁵.

For DC1, we simulate each CCD of the focal plane individually, we generate a single image with a 30-second exposure time to simplify the data handling. We omit instrumental effects and variability in the optical model across the focal plane. Since this is the first DESC data challenge, we want to ensure that we are able to generate and process the simplest cases, and then, building upon

this base, increasing the complexity and level of realism for future data challenges. Our sky brightness model is based off the Krisciunas & Schaefer (1991) model provided by OpSim, refined by the detailed wavelength dependence of the phenomenological model from Yoachim et al. (2016). The PSF model is a Gaussian for the system with a full-width half-maximum airmass dependence⁶LSST-20160 is not public..., this is done to mimic the degradation in the image quality due to, e.g., gravity load⁷. A Kolmogorov profile is used to model the atmosphere which is also airmass dependent⁸. The airmass, X , depends on the angular distance to the zenith, Z , as follows (Krisciunas & Schaefer 1991):

$$X = (1 - 0.96 \sin^2 Z)^{-0.5}. \quad (1)$$

imSim can generate three different types of objects: stars, which are modeled as PSF-like objects; galaxies, which are modeled as composite (bulge plus disk) Sérsic profiles (Sérsic 1963) using the parameters given by CatSim; and AGNs which are also modeled as point sources and, for simplicity, without any variability. Future versions of imSim will have the ability to generate more complex galaxy morphologies. The brightness for these sources is computed using the magnitudes from CatSim, which are converted to counts using the latest version of the LSST throughputs⁹. In DC1, we clip the objects at magnitude 10 in order to improve the computational efficiency.

The final products of this pipeline are FITS images with information about the observing conditions. We generated more than 200,000 images in total (including both the *dithered*, and *undithered* fields). The average time to simulate each CCD is ~ 4300 seconds and the total production time is $\sim 270,000$ CPU-hours.

5.2. Image processing

The outputs of these simulations are then processed using the LSST Science Pipelines (Ivezic et al. 2008; LSST Science Collaboration 2009; LSST Dark Energy Science Collaboration 2012; Bosch et al. 2018; Jurić et al. 2015) using version 13.0¹⁰, which we will refer to as the DM stack. The DM stack is an open-source, high-performance data processing and analysis system intended for use in optical and infrared survey data. The code can be found at dm.lsst.org and pipelines.lsst.io. A

⁴ <https://github.com/LSSTDESC/imSim>

⁵ <https://github.com/LSSTDESC/imSim/releases/tag/0.1.0>

⁶ From LSST-20160 eqn. (4.1) (RM: should link to this document / name it)

⁷ See LSE-30 <http://ls.st/lse-30> p.80

⁸ From LSST-20160 eqn. (4.2)

⁹ <https://github.com/lsst/throughputs>

¹⁰ <https://pipelines.lsst.io/releases/v13.0.html>

brief schematic of some of the steps in the image processing pipeline can be seen in Figure 1 as green squares. The raw, uncalibrated single exposures are used as inputs. The software performs the reduction, detection, deblending and measurement on individual visits. It then combines the single-visit images to produce the so-called coadds. This is done by computing a weighted average of resampled overlapping sensor images. For more information about this process, see Section 3.3 in Bosch et al. (2018). After assembling the coadd images, the DM stack performs measurements on them producing a catalog. The DM stack provides calibrated images and source catalogs for the individual visits and coadds stored in FITS files. In total, we detect and measure ~ 10.6 (9.7 for the undithered simulation) million objects with position, flux and shape information. We activated optional extensions for the pipeline to include CMODEL fluxes (see Bosch et al. (2018) for more details) and HSM shapes (Hirata & Seljak 2003; Mandelbaum et al. 2005). (JS: Ping Jim Chiang/Chris Walter/Tom G./Tony J.? to write something about the processing workflows for DC1.) An example coadd cutout is shown in Figure 4.

The reduction pipeline is essentially the same as for HSC (Bosch et al. 2018). This is going to allow us to use the HSC selection criteria as basis for our analysis, and can potentially enable direct comparisons between datasets for further validation.

(RM: Big picture comment: we don't talk at all about any of the software that drove these things - workflows for the DM processing etc. Is that sufficiently trivial that it doesn't need to be mentioned?)

6. OUTPUT CATALOGS AND VALIDATION

After being processed, the catalogs are stored at the National Energy Research Scientific Computing Center (NERSC)¹¹ and accessible to DESC collaborators. We generate `pandas dataframes` (McKinney 2010) and databases for each one of the coadds and input catalogs, providing collaborators some flexibility in data access for their own analyses. As mentioned earlier, the output catalogs contain 10.6 (9.7) million objects covering an area of ~ 43 deg². The catalogs include information about position, size, shape and magnitude for every object. They also include several flags that give information about the presence of interpolated/saturated pixels in the objects and whether or not these objects are close to the edge of a CCD.

6.1. LSST Science Requirements Document Key Performance Metrics and DESC Science Requirements Document

In order to check the level of realism and the accuracy of the processed catalogs we perform several quality assurance tests. These tests check two different aspects: the level of realism and consistency of the simulations, and the performance of the processing pipeline. The results in section are presented in the dithered simulation, however, unless stated, the procedures and behavior are similar for the undithered simulation.

Firstly, we perform some basic sanity checks and, after this, we process the output individual-visit catalogs through the LSST Project package `validate_drp`¹². The `validate_drp` package calculates the Key Performance Metrics (KPMs) from the LSST Science Requirements Document¹³ (Ivezić & the LSST Science Collaboration 2013), which we will refer to as the LSST-SRD. This document describes science-driven requirements for LSST data products and we use it as a guide to check the status of our end-to-end pipeline, with a special focus on the performance of the processing pipeline. In addition, we will also validate our dataset using some of the requirements at the DESC Science Requirement Document (The LSST Dark Energy Science Collaboration et al. 2018, DESC-SRD;), which we will refer to as DESC-SRD.

By design, our simulations do not satisfy some of the requirements in the LSST-SRD, such as the number of filters in the surveyed fields (Filter complement in Table 6.1) or the number of filters (Nfilt) used in a given night because we only simulate images in *r*-band. On the other hand, our images automatically meet some criteria due to the design choices, e.g., requirements about the pixels size since it is fixed. However, some requirements in both the LSST-SRD and the DESC-SRD cannot be tested with these single-band images (for example, KPMs involving colors) and we will ignore such tests in this work.

The results of **these checks are summarized in Table 6.1**. We proceed now to provide a description of these tests. More details about these tests and their motivation can be found in both the LSST-SRD (Ivezić & the LSST Science Collaboration 2013) and the DESC-SRD (The LSST Dark Energy Science Collaboration et al. 2018).

6.1.1. Astrometric performance

¹¹ <http://www.nersc.gov>

¹² [dmtn-0008.lsst.io, https://github.com/lsst/validate_drp](https://github.com/lsst/validate_drp)

¹³ <https://ls.st/LPM-17>

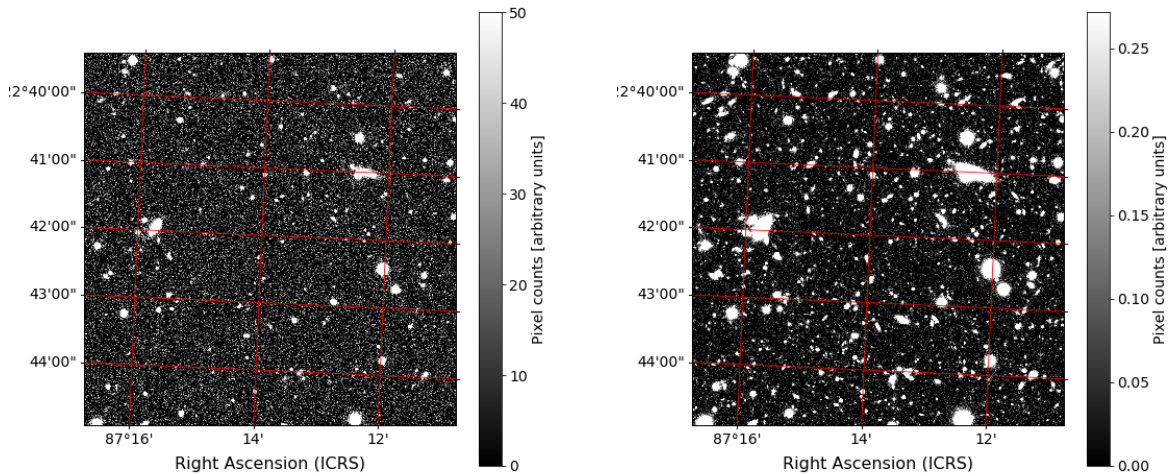


Figure 4. Example of a 1000×1000 pixel cutout from a calibrated exposure (left), i.e. background subtracted, reduced single-epoch image; and a full depth coadd (right). We can see the stark difference in the number of objects that are detectable by eye. (AS: Change this figure as follows: make a three/four panel figure showing: i) the true sky (basically imsim run with noise and psf, etc switched off); ii) a raw uncalibrated image of the sky; iii) the single calibrated exposure if sufficiently different – the scale should now how different units and you can draw line of const ra/dec to indicate astrometric calibration and iv) the 10yr coadd.) (JS: I don't have the clean imSim image before PSF convolution etc etc... The raw only has the background so I don't know if it's worth adding it)

We start by checking the absolute astrometry by comparing the input and output catalogs in Figure 5. This test helps us determine if the galaxies are generated in the correct position in the sky, and if our pipeline is able to accurately and precisely reconstruct the position of the simulated objects. A precise astrometric solution is necessary in order to compute unbiased clustering statistics, given that they mostly depend on the separation distance between objects. We find a mean (and median) deviation of 38 milliarcseconds between the input and measured positions. This is due to the fact that the corrections for proper motion were not present in the version 13.0 of the DM stack. However, this bias is still below the minimum specifications for the absolute astrometric performance of LSST as defined by the LSST-SRD, which is 100 milliarcseconds, and is driven mainly by orbital computations of solar system objects (Ivezić & the LSST Science Collaboration 2013).

We now focus on checking the astrometric repeatability in Figure 6. We check that the positions and distances between objects are consistent among different visits, this way, we will obtain consistent clustering results. We compute the RMS of the separation between pairs of stars separated by $D = 5, 20$ and 200 arcmin, obtaining $AM1=8$ mas, $AM2=4$ mas and $AM3=7$ mas. We check that no more than $AF1=AF2=20\%$ ($AF3=30\%$ in

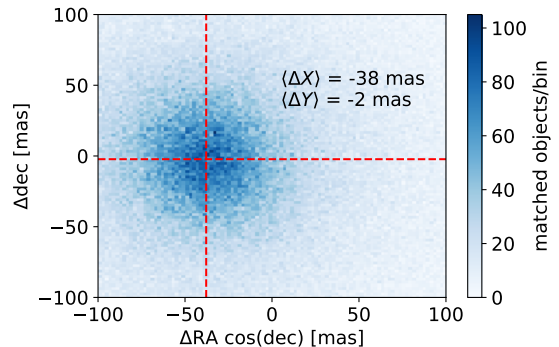


Figure 5. Astrometric residuals when comparing with the input catalogs for 500 randomly selected visits in the dithered simulation. We find a bias of 38 milliarcseconds in the measured right ascension due to uncorrected proper motion. We obtain similar results using the undithered simulation.

case of $D = 200$ arcmin) of these pairs deviate more than 40 mas (50 mas for $D = 200$ arcmin) from the median.

We find that our dataset pass all of these astrometric requirements, guaranteeing that the positions will be useful for clustering analyses.

6.1.2. Photometric performance

Accurate and consistent photometry is important in order to have a robust flux limited sample, and well behaved photometric redshifts. This is why we also val-

Table 1. Summary of Key Performance Metrics (KPMs) and DESC-SRD requirements that we check to test the quality of our end-to-end pipeline. The criteria that fixed due to design choices are shown in blue.

KPM/Requirement	Minimum	DC1	Passed (Y/N)	Section	Figure(s)
Filter complement	ugrizy	r	N		
Nfilters	3	1	N		
Nv1	184	184	Y		
pixSize (arcsec)	0.22	0.2	Y		
AA1 (milliarcsec)	100	20	Y	6.1.1	5
AM1 (milliarcsec)	20	8	Y		6
AF1 (%)	20	13	Y		6
AM2 (milliarcsec)	20	4	Y		6
AF2 (%)	20	9	Y		6
AM3 (milliarcsec)	30	7	Y		6
AF3 (%)	20	2	Y		6
PA1 (millimag)	8	6	Y	6.1.2	7
PF1 (%)	20	17	Y		7
PA3 (millimag)	15	0.06	Y	6.1.3	9
PF2 (%)	20	0	Y		9
PA6 (millimag)	20	17	Y		9
D1 (mag)	24.3	24.3	Y	6.1.4	10
Z1 (mag)	24.0	24.1	Y		10
DB1 (mag/r-band)	24.3	24.3	Y		10
Z2 (mag)	0.4	0.1	Y		
SE1	0.04	0.001	Y	6.1.5	11
SE2	0.1	0.002	Y		11
SR1 (arcsec)	0.80	0.64	Y		
SR2 (arcsec)	1.31	1.01	Y		
SR3 (arcsec)	1.81	1.79	Y		
TE1	3×10^{-5}	3×10^{-6}	Y		12
TE2	3×10^{-7}	9×10^{-8}	Y		12
WL4-Y10 (%)	0.1	0.05	Y		13

idate the photometric repeatability given by our processing pipeline by comparing the measured magnitude of bright ($\text{SNR} > 100$) unresolved objects across different visits. This test validates that the pipeline is reconstructing fluxes of objects consistently across epochs, but also, that different epochs are produced consistently in our image simulations. The photometric repeatability (PA1) is 6 mmag as measured by the scaled interquartile range¹⁴, and 15.8 mmag as measured with the RMS; these distributions are shown in Figure 7. The minimum specification is in the LSST-SRD is 8 mmag so

our dataset and pipeline behave as required. The significantly greater RMS is dominated by a tail of scattered objects, see Figure 8.

We can separately look at the reported photometric uncertainty from the processing. The Science Pipeline processing does not explicitly compare the reported photometric uncertainty to the per-epoch variation. Thus comparing the quoted uncertainties to the observed variation of stars with no intrinsic variation validates the consistency between the pipeline calculations and the model outlined in Ivezić et al. (2008, Eq. 4,5):

$$\sigma^2 = \sigma_{\text{sys}}^2 + \sigma_{\text{rand}}^2 \quad (2)$$

$$\sigma_{\text{rand}}^2 = (0.04 - \gamma) 10^{0.4(m-m_5)} + \gamma 10^{0.8(m-m_5)} \quad (3)$$

¹⁴ IQR \equiv 75th percentile minus 25th percentile, and then divided by the interquartile range of a Gaussian distribution with $\sigma = 1$

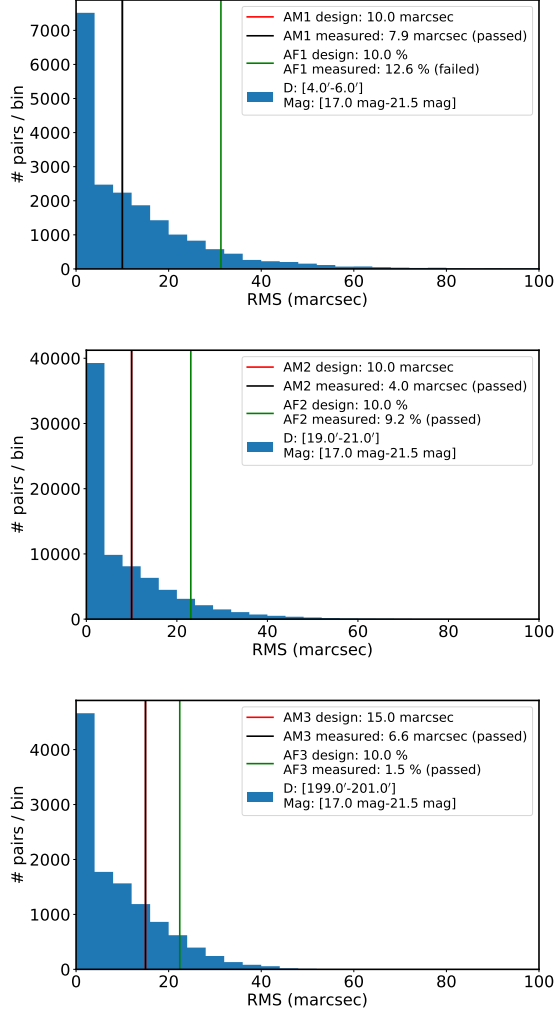


Figure 6. Astrometric repeatability: variation in distances measurement between pairs of stars at (top) 4-6', (middle) 19-21', (bottom) 199-201'. AM(1,2,3) are the distance measurements, while AF(1,2,3) are the fraction of pairs lying outside the a specified limit AD(1,2,3). The performance is excellent, with characteristic values all below the LSST-SRD levels.

where σ is in magnitudes, γ describes noise from the sky and detector electronics, and the 5- σ point-source depth is given by:

$$m_5 = C_m + 0.50(m_{\text{sky}} - 21 \text{ mag/arcsec}^2) + 2.5 \log_{10}(0.7''/\theta_{\text{eff}}) + 1.25 \log_{10}(t_{\text{vis}}/30 \text{ s}) - k_m(X - 1) \quad (4)$$

where C_m summarizes the throughput of the telescope optics and camera, m_{sky} is the sky brightness, θ_{eff} is the seeing, t_{vis} is the exposure time, k_m is the airmass coefficient, and X is the airmass.

Our fits find $m_5 = 24.2 \text{ mag/arcsec}^2$ and $\gamma = 0.038$ which are consistent with the Ivezić et al. (2008, Table

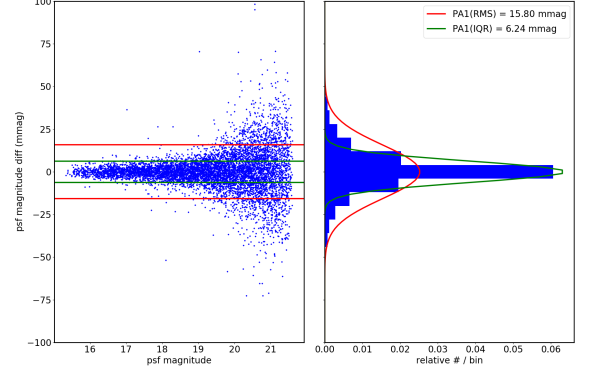


Figure 7. (left) The magnitude difference of pairs of measurements of stars across visits for stars with a typical SNR > 100 . (right) The histogram of these differences. The Gaussian root mean square (RMS) is shown in red while the interquartile range is shown in green. Note that the distribution is more peaked than a Gaussian. The interquartile range (6.2 mmag) is *smaller* than the Gaussian RMS (15.8 mmag). This means that we have extended tails but most objects have very accurate magnitude measurements.

2) values of $m_5 = 24.35 \text{ mag/arcsec}^2$, $\gamma = 0.039$. This demonstrates that we generate simulations of the telescope, detector, and sky in accordance with the Ivezić et al. (2008) estimates, as expected, since our inputs are based on these models.

Note that we find $\sigma_{\text{sys}} = 0 \text{ mmag}$, which is consistent with the idealized simulations we run for DC1 with no additional systematic sources of errors. We also check the fraction of photometric, bright unresolved sources that deviate more than 15 mmag, finding only PF1=17% over this threshold, in compliance with the requirements (PF1=20%).

6.1.3. Zeropoint uniformity

In order to obtain accurate photometric redshifts we need stable zeropoints across the sky, i.e., the error in the zeropoints, σ_{zp} fulfill the uniformity requirements specified in the LSST-SRD. More concretely, they should show an RMS lower than PA3=15 mmag and no more than PF2=20% of the images should have a deviation larger than 20 mmag. We randomly select 1000 sensor-visits and check the distribution of σ_{zp} , depicted in Figure 9. We find that the RMS of the distribution is 0.06 mmag fulfilling the requirements found in (Ivezić & the LSST Science Collaboration 2013). We also find that PF2=0% of the sensors have an error in the zeropoint that deviates from the median more than 20 mmag. Finally, we compare input and output magnitudes and compute the median difference between both finding it to be PA6=17 mmag and smaller than the maximum allowed in the LSST-SRD (PA6=20 mmag).

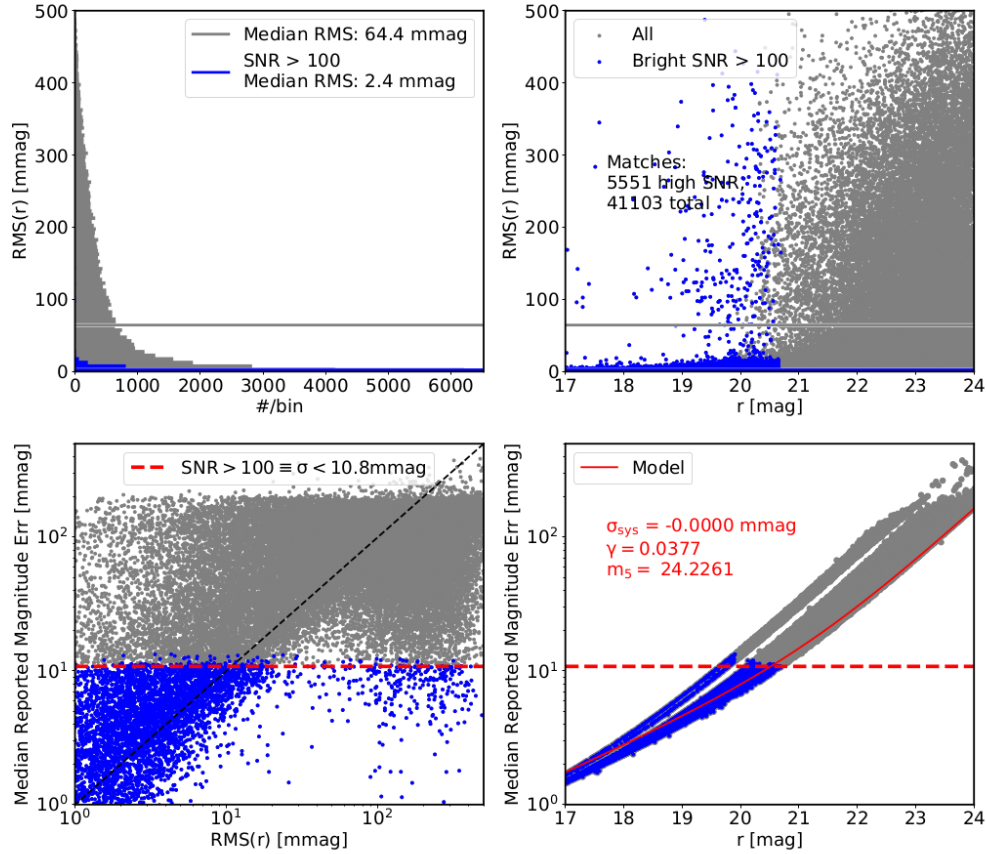


Figure 8. Photometric performance in more detail. (upper left) Horizontal histogram of the RMS distribution for each star for the full sample (grey) and “bright” sample (blue). The blue sample is the same as that shown in Figure 7. (upper right). Note that this RMS is the RMS of the measurements for a given star and so each entry in the histogram is one star. Thus, the median value here is 2.4 mmag, even though the median of all RMS measurements in Figure 7 was 15.8 mmag. This difference is due to a population of objects that have a poor repeatability. (upper right) The RMS distribution as a function of magnitude of the star. Here you can see the tight core of very repeatable measurements, along with a smaller population of objects with much higher scatter to their measurements. (lower left) Per-object median reported magnitude error by the pipeline versus the observed RMS over all of the measurements for an object. Our $SNR > 100$ cut is in the space of median reported error. (lower right) Per-object median pipeline-reported magnitude error versus object magnitude.

More details about how the matching between inputs and outputs was done in order to perform this test can be found in 7.1.

These tests show that the processing pipeline behaves as expected but also, that our images are generated with consistent zeropoints.

6.1.4. Depth requirements

In order to observe the object number density required to fulfill the science goals of LSST and DESC, the LSST-SRD sets some requirements in the minimum per-visit image depth for images with fiducial sky brightness of 21 mag/arcsec², exposure time of 30 s, airmass=1 and

fiducial seeing (FWHM) of 0.7 arcseconds. In order to mimic this we select the visits that fulfill the following criteria (Ivezić & the LSST Science Collaboration 2013):

- Altitude > 80 degrees.
- $0.68'' < \text{seeing (FWHM)} < 0.71''$
- Sky-brightness (in r -band) ≥ 21

We obtain a total of 520 sensor-visits fulfilling these criteria. We then compute the median 5- σ depth following the methodology described later in this work and compare with the predicted depth by OpSim. After this, we

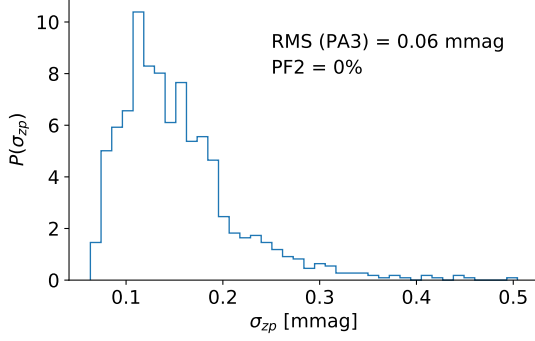


Figure 9. Normalized distribution of the error in the zero-point value, σ_{zp} for 1000 randomly selected sensor-visits. We see that the distribution is quite asymmetrical and find that the RMS (PA3) is 0.06 mmag, fulfilling the requirements in the LSST-SRD (15 mmag). We also find that only 0% (PF2) of the visits deviate from the median more than 20 mmag, in compliance with the LSST-SRD requirements.

check that the median of the depth distribution is larger than the minimum depth ($D1=24.3$ mag) as defined in the LSST-SRD. In order to have a uniform, well characterized sample, depth uniformity is already required. This is why we also check that no more than 20% of the visits have a depth lower than $Z1=24.0$ by computing the lower 20th percentile in the depth distribution. The results of these checks are depicted in Figure 10. We find that the median of the depth distribution in the selected visits, 24.297 ± 0.009 mag, is compatible (within a standard deviation of the mean) with the minimum depth set in the LSST-SRD. We find as well that the 20th percentile, $Z1=24.1$, is larger than the minimum value set in the LSST-SRD. We also check that in a given visit, the variation in the field-of-view is within the requirements. The LSST-SRD establishes that, in a representative visit with depth D1 no more than 20% of the field-of-view will be ($Z2$) 0.4 magnitudes brighter than the nominal (24.3). We select visit 2218486 since its median depth is 24.3. We find that the 20th percentile is 24.29 fulfilling the criteria. This test demonstrates that we generate images with the required depth and in concordance with our inputs.

6.1.5. PSF requirements

The LSST-SRD sets criteria regarding the maximum modulus of the PSF ellipticity, $|e|$, mostly driven by weak lensing analyses. We use the distortion definition for $|e|$ (Miralda-Escude 1991):

$$|e| = \frac{a^2 - b^2}{a^2 + b^2} \quad (5)$$

were a, b are the semi-major and semi-minor axes of the PSF. We test exposures with PSF-FWHM $\approx 0.69''$ and

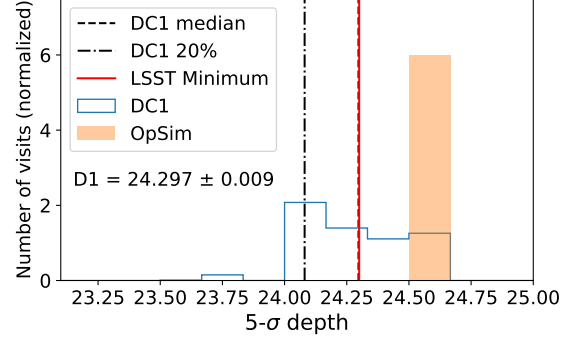


Figure 10. Measured depth in visits with Altitude > 80 deg., $0.68'' < \text{seeing} < 0.71''$ and Sky-brightness (in r -band) ≥ 21 (blue histogram) compared to the predicted depth by OpSim (solid orange histogram). The median of this distribution (dashed line) is very close to the LSST-SRD minimum depth $D1=24.3$ (red vertical line), the 20th percentile is also shown and we can appreciate that is larger than $Z1=24.0$ as established by the LSST-SRD.

no more than 10 degrees apart from zenith and check that the median ellipticity is no larger than 0.05 (SE1) and that no more than 10% of the images exceed 0.1 (SE2). Our analytic (and circularly-symmetric) PSF models should, by design, fulfill these criteria. However, we have to test if the reconstructed PSF also fulfills them. The PSF was reconstructed using the PSFEx (Bertin 2011) implementation in the LSST software stack. We tested this in the processed data by using the same 520 sensor-visits used to check the depth requirements described above. We checked the modulus of the PSF ellipticity in the position of the detected objects in these visits accumulating them in the histogram shown in Figure 11. We obtained that $SE1=0.001$ and $SE2=0.002$, below the maximum values allowed by the LSST-SRD. In future data challenges, we plan to increase the complexity and realism of the PSF to be closer to the maximum values set by these requirements.

We also check that, in these images, 85% of the point-like sources flux is contained within $0.80''$ or less (SR1), 95% within or less $1.31''$ (SR2) and 99% within $1.81''$ (SR3) or less. In our case we obtain $SR1=0.64''$, $SR2=1.01''$ and $SR3=1.79''$. This was done by calculating the radius at which the PSF was at its 85th, 95th and 99th percentiles.

For weak lensing analyses, the correct modeling of the PSF is crucial (Hirata et al. 2004) and both the LSST-SRD and the DESC-SRD contain explicit requirements about the PSF residuals. In particular, the LSST-SRD states that using the full survey data the auto- and cross-correlations (E_1, E_2, E_X) of the PSF residuals over an arbitrary field-of-view should be below TE1 (3×10^{-5})

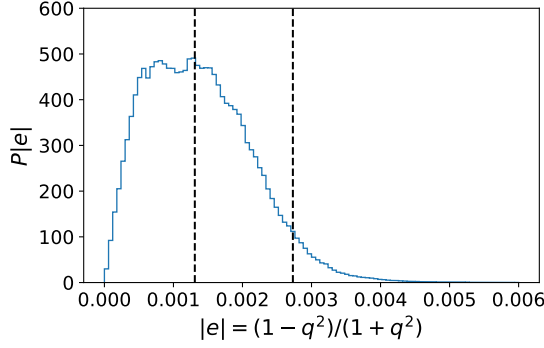


Figure 11. PSF ellipticity distribution accumulated for 520 sensor-visits measured at the position of detected objects. The median (SE1=0.001) and 90th (SE2=0.002) percentiles are shown as the dashed lines. Note that these values are an order of magnitude lower than the specifications by the LSST-SRD (SE1=0.05 and SE2=0.1, respectively).

for $\theta \leq 1$ arcmin, below TE2 (3×10^{-7}) for $\theta \geq 5$ arcmin and no more than 15% of the images will have median larger than TE3 (6×10^{-5}) for $\theta \leq 1$ arcmin and, TE4 (5×10^{-7}) for $\theta \geq 5$ arcmin.

To check these criteria we calculate E_1, E_2, E_X in using the definitions of the LSST-SRD:

$$e_1 = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad (6)$$

$$e_2 = \frac{2\sigma_{12}^2}{\sigma_1^2 + \sigma_2^2}, \quad (7)$$

$$E_1(\theta) = \langle \delta e_1^{(i)} \delta e_1^{(j)} \rangle, \quad (8)$$

$$E_2(\theta) = \langle \delta e_2^{(i)} \delta e_2^{(j)} \rangle, \quad (9)$$

$$E_X(\theta) = \langle \delta e_1^{(i)} \delta e_2^{(j)} \rangle, \quad (10)$$

With σ_1^2, σ_2^2 are the second order moments of a source along some set of perpendicular axes and σ_{12}^2 is the covariance, $\delta e_1, \delta e_2$ are the residuals, and the angle brackets indicate averaging over all pairs of stars i, j at a given angular separation θ .

In practice, we compute the PSF-corrected moments of high signal-to-noise ($\text{SNR} > 100$) stars across the field of view using **TreeCorr** (Jarvis et al. 2004). Our findings are shown in Figure 12. We can see that for E_1, E_2 , and E_X the TE1 criteria is fulfilled as well as TE2. We ignore TE3 and TE4 because we expect our images to fulfill these criteria automatically, given that we use the same analytic form for the PSF for all visits.

Finally, the DESC-SRD requires (WL4-Y10) that the systematic uncertainty in the PSF model defined using the trace of the second moment matrix should not exceed 0.1% for full-depth (Y10) DESC weak lensing analysis. We randomly select 3000 visits, obtain the input

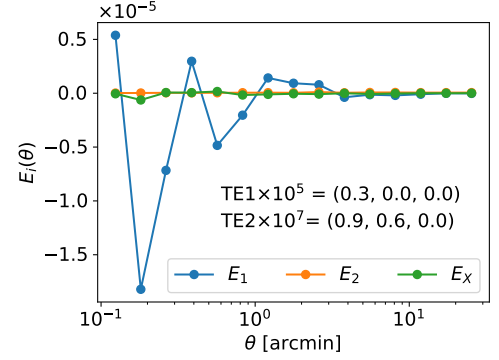


Figure 12. Auto and cross-correlation functions, E_1 (blue), E_2 (orange), E_X (green) of the PSF residuals as a function of the aperture angle θ . We see that our data fulfill the LSST-SRD requirements.

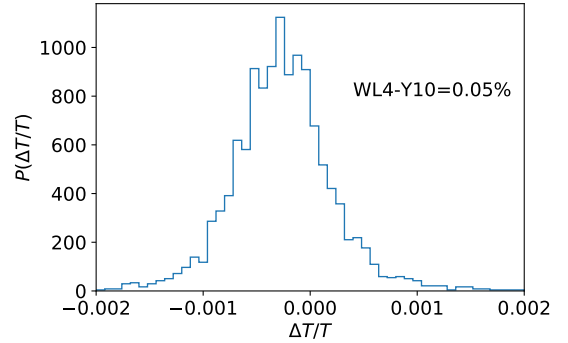


Figure 13. Normalized distribution of the relative difference in the trace of the second order moments, $\Delta T/T$ between input and output PSF. We see that the standard deviation of the distribution (WL4-Y10) is 0.05%, much complying with the requirement (0.1%).

and measured PSF, and measure the trace of the second order moments, T with **GalSim**. We then compute the relative difference, $\Delta T/T$, obtaining the results depicted in Figure 13. We find that our dataset shows that the standard deviation of the distribution is 0.05% lower than the requirement.

As seen in Table 6.1, our images and catalogs pass all the requirements by a good margin, demonstrating our ability to both generate and process high-quality data. We can also see that some of the criteria (e. g., number of filters used) are not passing the requirements due to design choices and do not affect the science analyses that can be performed with this dataset.

7. DATA SELECTION AND MASKING

In this section, we describe how we take advantage of the fact that we have full knowledge about the simulated sources in order to get a “clean” data sample, we also

describe the catalog mask and how we generate maps of different observational effects (seeing, sky-brightness, etc.) present in the simulation. Unless explicitly stated, the procedures and selections made in this section are performed in both the dithered and undithered simulations.

7.1. Matching inputs and outputs

Using end-to-end simulations, one can potentially trace each measured photon to its corresponding source and fully characterize the image generation and measurement processes. In practice, this is very difficult to do because, not just because of the large data overhead, but because the entire data reduction pipeline is built around pixelated images rather than tagged photon counts.

Nevertheless, the output catalog is a noisy representation of the underlying true catalog and we need to find a way to connect the two. The simplest way to connect two catalogs is by using the positions of the objects in the sky. This approach has been extensively used in the literature (de Ruiter et al. 1977; Benn 1983; Wolsencroft et al. 1986) and performs reasonably well when blending is low (i.e, when the amount of overlapping sources in the image are low). However, when blending is high, this approach might not be sufficient. Then, matching other quantities like flux, color and/or shape can become useful (Budavári & Szalay 2008; Budavri & Loredó 2015). For more details about challenges relating two different catalogs, we refer the reader to Budavri & Loredó (2015).

We compare two different matching strategies: positional matching, where we find the objects in the truth catalog closest to the detected objects, which we will refer to as *pure spatial matching* and will denote as S; and positional matching with magnitude matching, which we will refer to as *spatial+magnitude matching* and denote as S+M, where for each detected object we find objects from the input catalog that lie within a three pixel radius (0.6"). After this, we select the object that is closest in magnitude as long as the difference in magnitude is less than a certain threshold. In our case, we conservatively choose 1.0. Using this approach, if none of the neighbors fulfill these conditions, the detected source is considered unmatched. Note that in the case of pure spatial matching all sources are matched to an object in the true catalog.

In both cases, we build a `KDTree` (Pedregosa et al. 2011) using the positions of detected objects flagged with `detect.isPrimary=True` which ensures that the source has been fully deblended and was detected in the inner part of the coadd regions (see Bosch et al.

(2018) and Mandelbaum et al. (2018) for more details). In order to speed up the processing and to reduce the usage of computational resources we build the `KDTree` using output sources from 30 randomly selected coadd regions (patches) in the dithered simulation (the undithered simulation yields the same results and conclusions) containing 975,605 detected sources fulfilling the aforementioned condition (we will refer to these as primary outputs or primary detected sources). Using this sample, we find that 95% of these sources are matched using the spatial+magnitude matching (by construction all of them are matched using the pure spatial matching approach). An interesting metric to check is the fraction of primary detected sources that have been matched to the same exact object in the input catalog, which we will denote as f_{multi} . This is an unusual occurrence, however, imagine the case where an spurious fluctuation has been marked as a source, this fluctuation is a primary detected source but has no counterpart in the input catalog. Another example are bright objects that have been shredded and are detected as several fainter sources. We find that these kind of matches are ~ 100 times more likely to happen using the pure spatial matching ($f_{\text{multi}} = 3 \times 10^{-3}$) than in the spatial+magnitude matching ($f_{\text{multi}} = 2 \times 10^{-5}$). This is expected because, in the cases where the primary detected source is a random fluctuation or a shredded source, it is unlikely that the measured flux is close to the flux of a neighboring source thus producing an unmatched source in the case of using spatial+magnitude matching. We also find that both approaches select the same best matching source in the input catalog for only 68% of the primary detected sources for which we found a suitable match. These differences can be due to several reasons, e.g. objects with a poorly determined centroid positions that are close to other objects in the input catalog (remember that the input catalog contains objects up to $r = 28$), objects with poorly determined fluxes (low SNR), etc.

We compare as well the astrometric and photometric residuals using both approaches in Figure 14. We can see that the median astrometric residuals are similar in both cases and that the search radius (3 pixels=600 mas) for the spatial+magnitude matching is large enough to map the overall distribution. On the other hand, the resulting median photometric residual seems strongly biased in the case of using just pure spatial matching, and the scatter is very large as shown by the error-bars. However, in the case of spatial+magnitude matching, the median residuals and their error-bars are smaller, as expected given the maximum magnitude threshold. We can see that in this case the biases are still significant. This can be mitigated using a smaller toler-

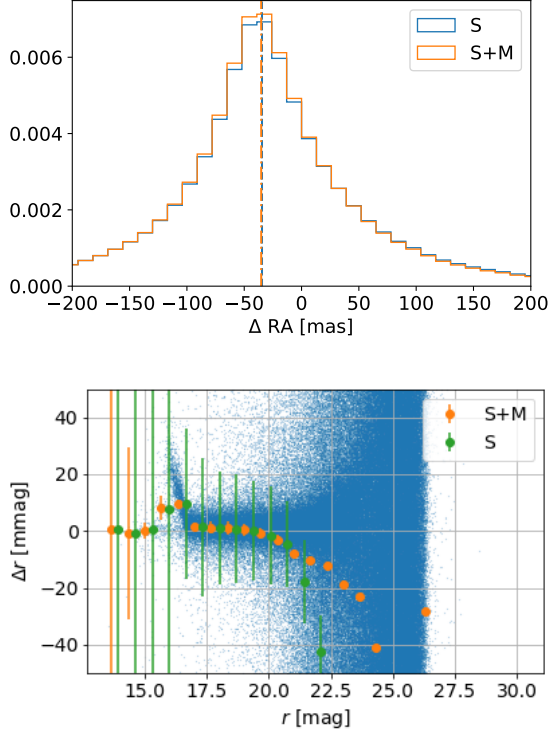


Figure 14. Top: Normalized distribution of measured astrometric residuals (in the RA axis but similar results are found for the Dec. axis) for pure spatial matching (blue) and spatial+magnitude matching (orange), the median values of each distribution are marked by dashed lines and dashed-dotted lines respectively. Bottom: Median per-bin photometric residual as a function of measured magnitude for pure spatial matching (green), and spatial+magnitude matching (orange) using 30 magnitude bins between r -band magnitude 10 and 30. We also show the individual residuals for the matched objects using spatial+magnitude matching (blue dots).

ance in magnitude (for example using 0.5 mag instead of 1 mag) but that implies an overall reduction of the number of matched primary detected sources. Finally, observing the magnitude difference between inputs and outputs of individual matched sources using the spatial+magnitude technique, we can see that this distribution is centered around zero, except for very bright sources ($r < 17$), where saturation prevents us from having an accurate determination of the fluxes.

Different matching techniques have different potential applications and strengths [Budavri & Loredo \(2015\)](#). In our case, we want to use these matching techniques to provide a clean (flux-limited) sample to perform two-point clustering analyses. Given that magnitude precision and accuracy will be important for our sample selection, the **spatial+magnitude** matching technique will be sufficient to clean the sample from artifacts and

poorly measured sources. However, more complicated matching techniques can become necessary for other use cases.

7.2. Sample selection

In this subsection we are going to use the spatial+magnitude technique to identify flags or thresholds in variables that may allow us to get a clean sample for clustering. In principle, given that the LSST DM software stack is essentially the same as the HSC reduction pipeline used in [Mandelbaum et al. \(2018\)](#), we could potentially perform similar cuts. However, note that we are working in r -band only so some of the required cuts cannot be performed. In addition to this, some variables, such as the so-called blendedness parameter, used in [Mandelbaum et al. \(2018\)](#), are not available in the version of the LSST DM software stack that we ran to process the data (v13.0). As a consequence, we propose our own selection cuts, although we follow the criteria in [Mandelbaum et al. \(2018\)](#) as guidance.

The methodology to perform the selections is simple: we check the primary detected sources that have no match using the spatial+magnitude technique and we compute the fraction of objects that are flagged, $f_{u,i} = N_{\text{flag}_i, \text{unmatched}} / N_{\text{total, unmatched}}$, and compare it to the corresponding fraction of flagged matched primary detected sources, $f_{m,i} = N_{\text{flag}_i, \text{matched}} / N_{\text{total, matched}}$, for each one of the flags, flag_i , in the catalog. If the ratio $f_{u,i} / f_{m,i}$ is larger than 50 for a particular flag and $f_{m,i} < 0.01$, i.e., less than 1% of the matched primary sources have that flag, it means that the presence of that flag is a good indicator of problematic sources. Thus, we eliminate the sources with those flags. We also repeat the same procedure looking for the absence of a certain flag or whether a quantity is frequently measured as not a number, NaN, and check for redundant cuts.

The results of the selection process are shown in [Figure 15](#). In this Figure we can see that most of the flags shown have a larger impact in measured sources that have no match in the input catalog (unmatched). We notice that some of the flags are very efficient distinguishing unmatched from matched objects. For example, `base.ClassificationExtendedness_flag = True`, which means that there was a failure at the time of deciding whether a source was extended or point-like, eliminates more than 30% of the objects with no match, while barely affecting the matched objects. There are other three flags pretty efficient to filter out unmatched objects but, in case of using them, we would lose $\approx 50\%$ of our sample. These include `modelfit_CModel.flags_smallShape==False` which means that the initial parameter guess did not result in a neg-

ative radius (if `True` the initial guess for the radius would be negative). Intuitively, we expect this flag to be `False` for well-behaved objects and thus, we should not use this for our selection. Another case where the ratio of flagged unmatched (and matched) objects is very large is `modelfit_CModel_flags_region_usedFootprintArea==True`, which means that the pixel region for the initial fit was defined by the area of the footprint. This flag, again, is not necessarily indicative of problems with the measured object. Finally, we see that `modelfit_CModel_flags_region_usedPsfArea==False` also affects a very large fraction of unmatched and matched objects. These objects are such that the pixel region for the initial fit was not set to a fixed factor of the PSF area, which is not indicative of any problems with the source. A summary of the effects of these cuts can be found in Table 2.

As a result, we are going to **eliminate** from our selection all the sources that fulfill at least one of the following conditions:

- `detect_isPrimary = False`. As discussed earlier, this means that the source has not been fully deblended or is outside of the inner region in a coadd.
- `base_NaiveCentroid_flag = True`. This means that there is a general failure during the source measurement.
- `base_SdssShape_flag_psf = True`. This means that there is a failure in measuring the PSF model shape in that position.
- `ext_shapeHSM_HsmSourceMoments_flag_not_contained = True`. This means that the center of the source is not contained in its footprint bounding box.
- `modelfit_DoubleShapeletPsfApprox_flag = True`. This means that there is a general failure while performing the double-shapelet approximation to the PSF model in the position of this source (see Appendix 2 in Bosch et al. (2018) for more details).
- `base_PixelFlags_flag_interpolated = True`. This means that there are interpolated pixels in the source’s footprint.
- `base_PixelFlags_flag_interpolatedCenter = True`. This means that the center of a source is interpolated.
- `base_PixelFlags_flag_saturatedCenter = True`. This means that the center of a source is saturated.

Cuts	Matched [%]	Matched kept [%]
Primary detected	95.2	100
Clean sample	96.5	99.7
All flags in Figure 15	98.7	32.3

Table 2. Summary of selection efficiency and purity for the different flag cuts that we explored in Figure 15.

- `base_ClassificationExtendedness_flag = True`. This means that there is a general failure when using the extendedness classifier.
- `modelfit_CModel_flags_region_usedInitialEllipseMin = True`. This means that the pixel region for the final model fit is set to the minimum bound used in the initial fit.
- `base_SdssShape_x/y = NaN`. This means that the centroid position (either in the x or y axes) is measured as NaN.
- `base_SdssCentroid_x/yErr = NaN`. This means that the error in the centroid position (either in the x or y axis) is measured as NaN.

After these cuts we keep 8.25 million objects in the dithered catalog and 7.51 million objects in the undithered catalog. We will refer to this sample as the *clean sample*. After the cuts, the ratio of unmatched objects decreases by $\approx 64\%$ from 5.6% (of the catalog before cuts) to 3.6% (of the clean sample), while we retain 99.7% of the matched objects. In DC1 we are somehow limited by the fact of having only information from *r*-band. The addition of other imaging bands will provide additional information to further clean the sample (e.g., by performing selection cuts in color-color diagrams).

We now focus on how many of these objects are matched as a function of magnitude and signal-to-noise ratio. In Figure 16, we can see that the fraction of unmatched objects grows very quickly after $r = 26$, and for $\text{SNR} < 6$. Therefore, $r < 26$ and/or $\text{SNR} > 6$ look like sensible selection criteria to ensure good quality data. In particular, we select our *final sample* using the following criteria:

- `base_ClassificationExtendedness_value=1`.
- $17 \leq r_mag_CModel \leq 25.5$.

These selection cuts ensure a low fraction of unmatched objects and also, as we will justify in following sections, good purity of our galaxy sample (low stellar contamination).

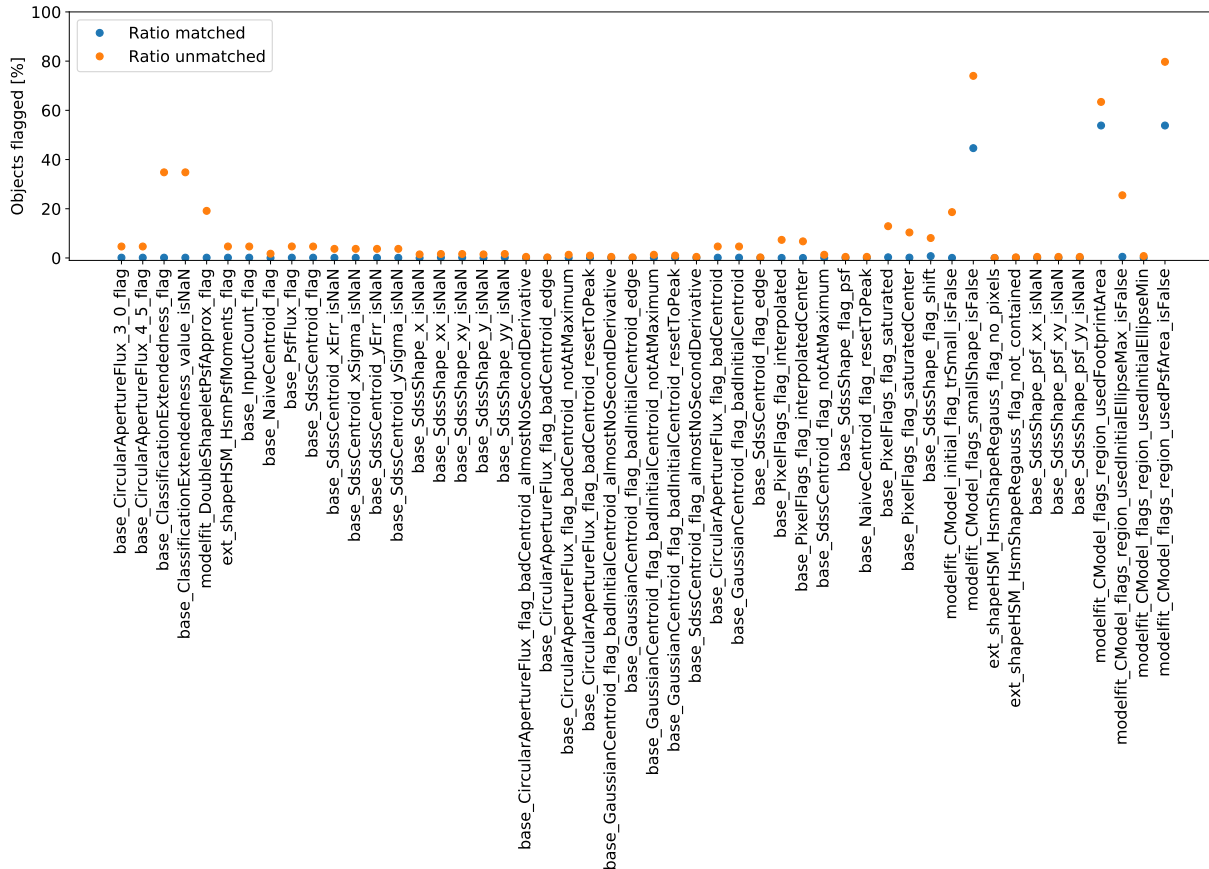


Figure 15. Percentage of flagged objects with a match in the (solid blue circles) and flagged objects with no match in the truth catalog (orange solid circles) as a function of each flag. We show only flags where there is either a large fraction of objects with no match compared to objects with a match or where more than 50% of the objects have been flagged and the fraction of unmatched objects is larger than the ratio of matched objects.

7.3. Star/galaxy classification

For weak lensing and clustering analyses, it is important to have a pure galaxy sample, and a good control over the fraction of stars that are classified as galaxies. Our pipeline includes the variable `base_ClassificationExtendedness_value` (see Bosch et al. (2018) for more details) , which we will refer to as extendedness, which can be used as a proxy to separate stars from galaxies as showed by Mandelbaum et al. (2018) and Bosch et al. (2018). In this work we say that an object has been classified as a galaxy if extendedness=1, and that the object has been classified as a star if extendedness=0. Now we are going to check how extendedness performs as star/galaxy classifier in DC1. In order to do so, we use the *clean sample* with $16 \leq r \leq 26$ in the dithered field, although we find similar results using the undithered field, and match it to our input catalog. After that, we count the number of objects matched to stars that have been classified as stars, as a function of magnitude, and divide this number by the total number of objects matched to stars (which we will mark as Star

→ Star). We do the same with the number of objects matched to stars that have been classified as galaxies, and divide this number by the total number of objects matched to galaxies (Star → Galaxy); the number of objects matched to galaxies that have been classified as galaxies divided by the total number of objects matched to galaxies (Galaxy → Galaxy); and finally, the total number of objects matched to galaxies that have been classified as stars, divided by the total number of objects matched to stars (Galaxy → Star). This way, we know the total stellar (or galaxy) contamination as a function of measured magnitude. The results are depicted in Figure 17. We see that at the brighter end ($r \leq 17$), there are a lot of stars that get classified as galaxies, this is due to two reasons: First, the ratio number of stars to number of galaxies at those magnitudes is larger than at fainter magnitudes, since stars and galaxies have different magnitude distributions. Second, problems related to saturation make difficult to use extendedness as a proxy for distinguishing stars and galaxies. On the other hand, we see that at the fainter end ($r \approx 26$), the purity

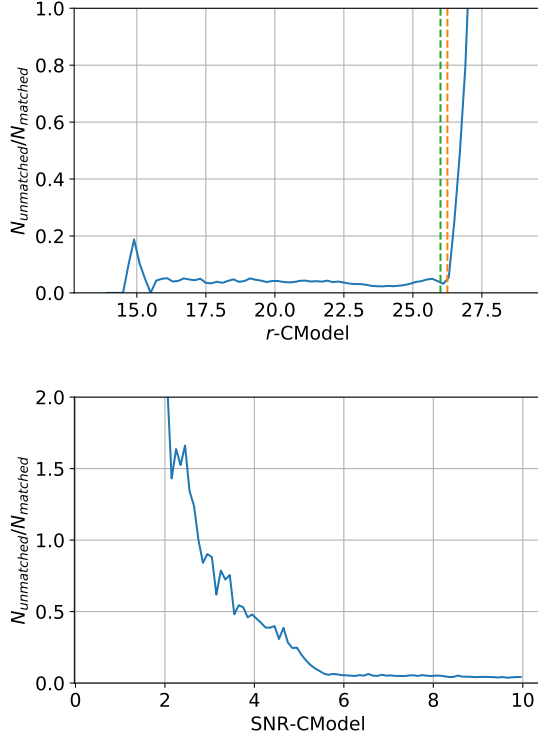


Figure 16. Top: Ratio of unmatched to matched primary detected sources after selection cuts as a function of magnitude. The dashed vertical lines show the median depth for the dithered (orange) and undithered (green) fields. Bottom: Ratio of unmatched to matched primary detected sources after selection cuts as a function of SNR.

of the galaxy sample using the extendedness classifier starts to decrease, getting as low as 90% for the last bin in our analysis, and an overall $f_{\text{star}} = 2.8\%$ stellar contamination in the *clean sample* after using extendedness=1. This number gets reduced to $f_{\text{star}} = 1.4\%$ if we only consider galaxies with $r < 25.5$, and $f_{\text{star}} = 0.7\%$ for $r < 25$. Note that these numbers are larger than those presented in Mandelbaum et al. (2018). This is mostly due to the fact that our PSF is larger, making the extendedness classifier perform a little bit worse, and that we do not include any cuts in resolution. However, this level of stellar contamination is acceptable for the purposes in this work.

7.4. Depth maps and footprint masking

In order to estimate the depth in the coadd catalogs we generate a flat-sky map, with resolution of 1.74 arcminutes, containing the detected primary sources in the *clean sample* from the previous subsection. This resolution ensures having an accurate estimate of the power-spectra up to $\ell \sim 6000$, where the power-spectra will be mostly dominated by shot-noise. Then, for each cell in

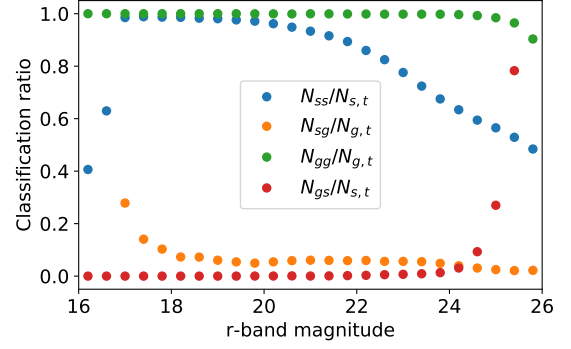


Figure 17. Performance of the `base-ClassificationExtendedness_value` as star-galaxy classifier as a function of magnitude. We show the ratio of the number of stars classified as stars N_{ss} and true stars $N_{s,t}$ (blue), the ratio of the number of stars classified as galaxies N_{sg} and the number of true galaxies $N_{g,t}$ (orange), the ratio of the number of galaxies classified as galaxies N_{gg} and the number of true galaxies $N_{g,t}$, and the ratio of the number of galaxies classified as stars N_{gs} and the number of true stars $N_{s,t}$. We see that the purity of the galaxy sample is $> 90\%$ in the magnitude range that we analyze.

the map, we compute the median SNR of the objects in the cell as a function of the magnitude, and use the magnitude at which SNR is closest to 5.

These maps are shown in Figure 18. We can see that the dithered simulation is indeed very uniform ($> 50\%$ of its footprint lie in the same exact depth bin) showing the success of the dither strategy. However, we can see that in the undithered simulation, as expected, there are zones with a higher depth (the overlap of the pointing positions) and a reduced median depth.

We also check the depth by checking the detection efficiency of stars as a function of magnitude. To do so, we use the stars in the input catalog and select those that lie within the simulated footprint, after this, we match them using the spatial+magnitude technique described in previous sections. Finally, we compute the number of detected objects in the *clean sample* matched to stars divided by the number of stars in the input catalog as a function of the magnitude of the matched object in the input catalog. The results can be seen in Figure 19. We see that there is a high efficiency $> 80\%$ up to $r \approx 25.8$.

Given the results in the two previous subsections and this subsection, we decide to use only the galaxies that lie in cells with limiting magnitude $r \geq 25.5$ and that have been visited, at least, 92 times, which corresponds to 50% of the nominal full-depth number of visits (Ivezic et al. 2008). On top of that, we select those objects with magnitude $17 \leq r \leq 25.5$. This cut ensures high com-

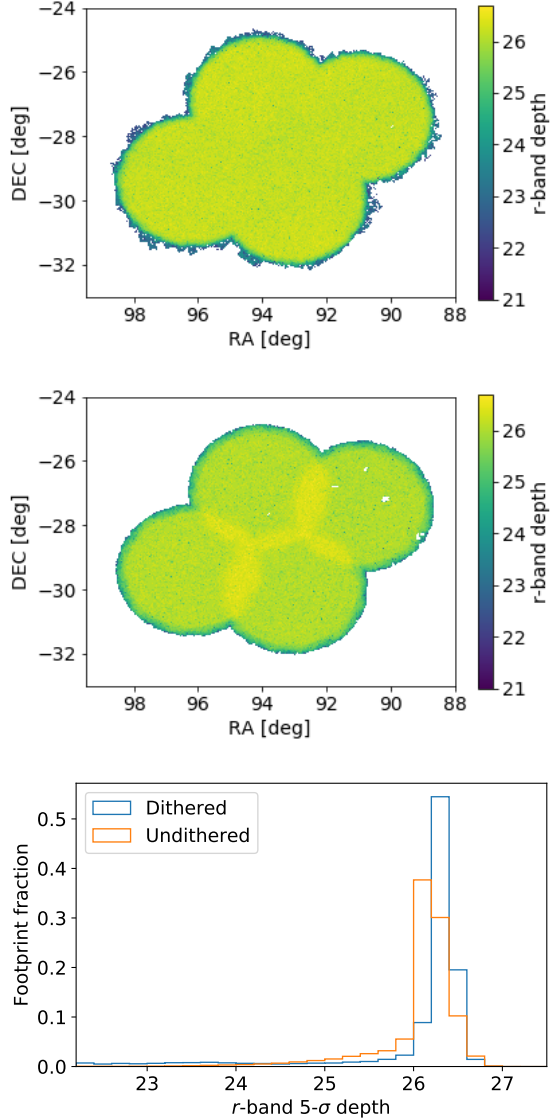


Figure 18. 5- σ depth maps for the dithered (top) and undithered (middle) fields. There is an increased depth in the overlapping parts of the pointings in the undithered field but the median depth is lower. The maps have a resolution of 1.74 arcmin. We also show the 1D distribution of depth (bottom) for both fields for easier comparison.

pleteness ($> 85\%$) and it allows us to eliminate most of the artifacts in the sample. In addition it results into a low stellar contamination ($f_{star} \approx 1.4\%$). After these cuts and selecting objects with extendedness=1, we obtain 5.0 and 4.6 million objects for the dithered and undithered fields respectively. This selection cut however, does not change the fraction of unmatched objects explored in the previous subsection.

7.5. Bright object masking

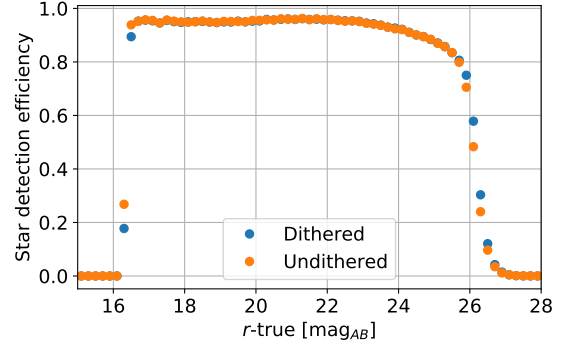


Figure 19. Ratio of number of detected objects matched to stars in the *clean sample* and number of stars in the input catalog as a function of magnitude for the dithered (blue) and undithered (orange).

Bright objects produce significant effects in the image that affect the detection and measurement of neighboring objects. Some examples of these effects include saturation, large diffraction spikes (not included in our simulations), obscuration of neighboring sources, etc. Thus, masking a region around these sources creates a more complicated footprint but greatly simplifies the analysis of systematic effects. In order to avoid possible biases by masking bright galaxies we will only analyze the impact of bright stars in the nearby detected objects. In the following, when we refer to bright object masking we are actually referring to bright star masking.

In order to evaluate the effects, we follow the procedure described in [Coupon et al. \(2018\)](#). Using the position of bright objects classified as stars (`base_ClassificationExtendedness_value==0`), that lie within the considered footprint, and with input magnitudes in the range $m_1 < r < m_2$ we count all objects from the final sample in a given radius θ , and compute the mean of the set, $N_{neighbors}$. We repeat this for different radii and magnitude ranges ($r < 17$; $17 \leq r < 18$; $18 \leq r < 20$; $20 \leq r < 22$). Finally, we repeat this process for all stars in the input catalog in the footprint and compute $N_{neighbors,tot}$ and compute the ratio $y = N_{neighbors}/N_{neighbors,tot}$. Following [Coupon et al. \(2018\)](#) we use the point where the density reaches 95% as masking radius. After this, we compare to equation (1) in [Mandelbaum et al. \(2018\)](#):

$$r_{mask,HSC} = 200 \times 10^{0.25(7-m_{B*})} + 12 \times 10^{0.05(16-m_{B*})} \quad (11)$$

Where m_{B*} is the measured magnitude of the bright star to mask. Note that this prescription is specific for HSC. However, the LSST instrument and observing conditions are different. This is why we decided to rescale $r_{mask,HSC}$ by the ratio of mean seeing in LSST

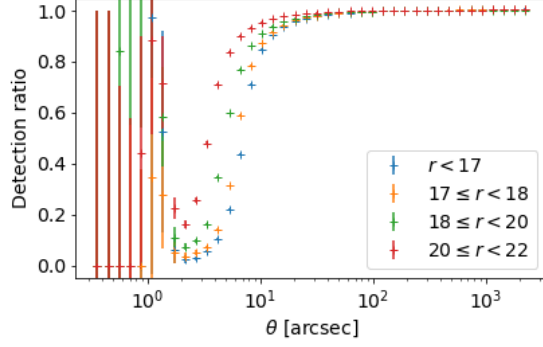


Figure 20. Ratio of the median number of primary detected objects neighboring a star in a certain magnitude range in the input catalog to the median number of objects detected any star in the input catalog, as a function of the distance to the star θ . Different symbols and colors represent different magnitude ranges for the stars in the input catalog considered. The shadowed regions represent the best fit masking radius, $r_{mask,fit}$ with their uncertainties and the vertical lines correspond to $r_{mask,LST}$ evaluated at the mean magnitude in each bin.

and HSC.

$$r_{mask,LST} = r_{mask,HSC} \frac{1.04''}{0.58''} \quad (12)$$

Where $1.04''$ is the mean seeing in DC1, and $0.58''$ is the mean seeing reported in Mandelbaum et al. (2018). We obtain the results depicted in Figure 20. We find that the masking radius, $r_{mask,LST}$, using the rescaled prescription set by Mandelbaum et al. (2018) is a pretty good approximation for our data. In particular, we find that the ratio $r_{mask,fit}/r_{mask,LST} \approx 1.14$. This 14% disagreement can be related to the fact that we chose a very simple rescaling based on the mean seeing and due to our simplistic PSF model. As a result, we decide to use $r_{mask,DC1} = 1.14r_{mask,LST}$ around stars with $r < 22$. We generate a high-resolution mask using flat-sky maps of $\approx 6.4''$ resolution, then we downsample this map to a resolution of ≈ 2 arcmin and eliminate objects that lie within pixels that have more than 75% of their area masked. This results into an area loss of $\approx 13\%$. The resulting map can be seen in Figure 21.

7.6. Blending

As previously mentioned, our output catalogs do not include any estimation of overlap between sources, or *blendedness* (Bosch et al. 2018). Highly-blended objects are more likely to have biased estimations of the centroid position, shape, and fluxes. This can lead to overall biases in the estimated photometric redshifts and cosmological parameters. Given that the mean seeing in

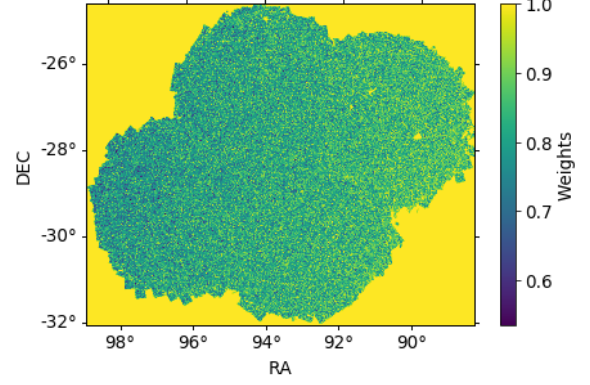


Figure 21. Bright object (bright star) map. This map shows the unmasked fraction in each pixel. We use a high-resolution ($\approx 6.4''$) map to mask around bright stars ($r < 22$) and then, we downsample the map to a lower resolution ($\approx 1.7''$) and get rid of the pixels where the masked area by bright stars is higher than 75% of the pixel.

DC1 is larger than in HSC, the impact of blended objects is likely to be larger. In the case of Mandelbaum et al. (2018), the cut in the blendedness parameter affects only 1% of the objects; we expect this number to be larger in our case. Using specialized image simulations from Sánchez et al. (in prep.) with a seeing similar to the seeing in DC1 ($1.04''$), in r -band, we find that if we select objects with $r < 25.5$ and $SNR \geq 1$, the fraction of objects with blendedness $> 10^{-0.375}$ is $\approx 6.3\%$. If we raise the minimum SNR threshold to 6, this fraction is lowered to $\approx 2.6\%$. This means that our sample will have a fraction of these objects anywhere in the range (2.6% - 6.3%) but closer to 2.6% since the fraction of objects with $SNR \leq 6$ is $\approx 0.3\%$. In any case, we do not expect that the inclusion of these objects in our two point measurements will affect the range of scales that we are going to consider in this work. In order to quantify this, a careful study of the impact of blending in clustering measurements is necessary but this is out of the scope of this work.

8. TWO-POINT CLUSTERING RESULTS

In this section, we analyze the two point clustering statistics for both the dithered and undithered catalogs in harmonic space and check the consistency between the input and measured observables. We also analyze the impact of different observing conditions in two-point statistics.

In order to measure the angular power spectrum, we use NaMaster¹⁵ (Alonso et al. 2019). NaMaster allows us to compute the cross-power-spectra of spin-0 and spin-

¹⁵ <https://github.com/LSSTDESC/NaMaster>

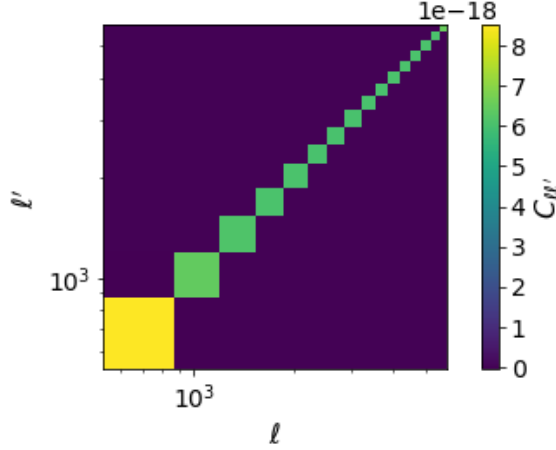


Figure 22. Estimated Gaussian covariance using NaMaster in the range of scales considered in our analysis.

2 maps, with an arbitrary mask and number of contaminants using the pseudo- C_ℓ formalism (Peebles 1973; Hivon et al. 2002; Elsner et al. 2017). In this case, we focus on the angular power spectrum of the density contrast map, δ . Given the lack of photometric redshift measurements in our data, we decided to analyze full *final sample*. We check the range $0 < \ell < 6000$ in band-powers of $\Delta\ell = 352$ and ignore the first bin due to lack of sensitivity at that range of scales ($\ell = 178 \approx 2$ degrees). The choice of $\Delta\ell$ was done to have ~ 3 times more bins than the optimal $\Delta\ell_{opt} = 1/f_{sky}$, with f_{sky} being the fraction in the footprint. However, we are going to only consider multipoles such that $\ell \geq 1/f_{sky} \sim 1192$. We estimate the covariance matrix using two different approaches: On the one hand, we compute the Gaussian covariance with NaMaster obtaining the results in Figure 22. On the other hand, we compute the jackknife covariance matrix by doing 155 equal area regions in our footprint, and correct using the Hartlap factor (Hartlap et al. 2007). We find both approaches to give similar results and use the results from the jackknife computation.

The DC1 simulations also allow us to study the effect of different observational effects in the two-point clustering statistics and how the dithering strategies work to mitigate these effects. We are going to consider the following effects:

- **Extinction:** The CatSim catalog provides the value for the magnitudes corrected for extinction using the map from Schlegel et al. (1998), which we refer to as the SFD map.
- **Stellar contamination:** In this case, we build a flat-sky map with all stars in the input catalog.

- **Sky-background/Sky-brightness:** We use the observed background level in each exposure and assign that value to the pixels in the flat-sky map that lie within that exposure. After this we calculate the mean value in each pixel to build the map with the same resolution as the mask (≈ 2 arcmin) which we deproject (Alonso et al. 2019).
- **Sky-noise:** We use the observed noise background level in each exposure and proceed as in the previous case to build a map.
- **Seeing:** We proceed as before and use the observed seeing in each exposure and build a map.
- **Number of visits:** We count the number of exposures overlapping with each pixel of our flat-sky maps.

These maps are shown in Figure 23 and Figure 24. We see that the spatial distribution of the different observing conditions are very different between the two simulations, even though the ranges in each of the observing conditions are very similar.

We use the mode deprojection from NaMaster (Alonso et al. 2019) to correct for the potential contamination of the maps described above. Mode deprojection, assumes that there is a linear dependency between the observed number density of galaxies and the contaminants. We checked that this is a good approximation for the different observing conditions. In addition, we compute the theoretical prediction for the power-spectra with CCL (Chisari et al. 2019):

$$C_\ell^{TH} = \frac{2}{\pi} \int dz \left(\frac{dn(z)}{dz} \right)^2 b^2(z) \int dk k^2 P(k, z) j_\ell^2(kr(z)) \quad (13)$$

where $P(k, z)$ is the power spectrum, $b(z)$ is the bias and $\frac{dn}{dz}$ is the number density as a function of redshift. We use the Millenium cosmological parameters (Springel et al. 2005) ($\Omega_m = 0.25, \Omega_b = 0.045, \Omega_\Lambda = 0.75, n = 1, \sigma_8 = 0.9, h = 0.73$), and the dn/dz built by using the true redshifts in galaxies matched in the input catalog. We use a bias, $b(z) = b_0/D_+(z)$, inversely proportional to the linear growth factor, $D_+(z)$, calculated using CCL and defined as in Chisari et al. (2019) and check that the qualitatively speaking, there is a good agreement between the measurements and the predictions. However, we do not expect to be able to fully describe the measured power-spectra, given the highly-nonlinear nature of the scales considered in our analysis.

In addition, we see that the overall impact of the systematics is smaller than 30% of the statistical uncertainty for $\ell > 1/f_{sky}$. We also see that the observing

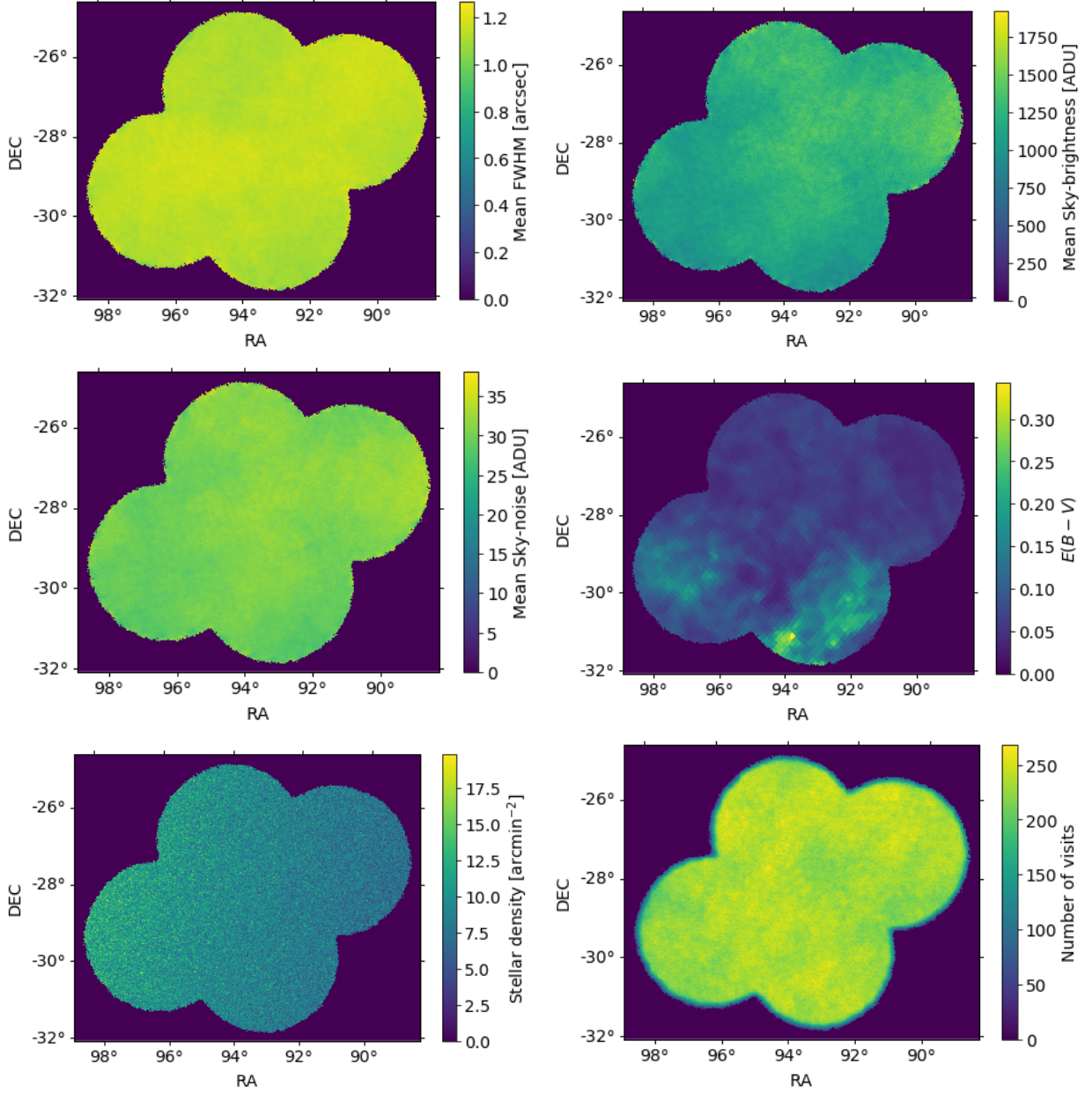


Figure 23. Maps showing the different foregrounds considered in our analysis of the dithered field. From top left to bottom right: Mean PSF FWHM, mean sky-brightness, mean sky-noise, mean extinction, stellar density and number of visits in each pixel in the flat-sky maps with the same resolution as the depth maps in Figure 18. We only show their values in the regions where the $5\text{-}\sigma$ r -band depth is larger than 25.5.

conditions similarly affect both dithered and undithered simulations, not finding any statistically significant difference between the correction for the dithered and undithered simulations. This is a consequence of several factors, including the conservative cuts that we impose on our data to ensure well-behaved clustering statistics; that we only deproject using the mean value for the different observing conditions; and the lack of effects, such as vignetting, present in real images. For example, vignetting would affect the number of detected objects

close to the edges of the focal plane in the undithered simulation, reducing the uniformity of the survey. However, this effect would be uniform across the footprint in the dithered case. In addition to this, if we decided to push further our data by going deeper, the systematic lack of uniformity of the undithered field would start to play a large role in the impact of the foregrounds. We can also see that, in the ℓ range that we consider for our analysis, the presence of bright objects – in particular bright stars – is the dominant systematic effect. In

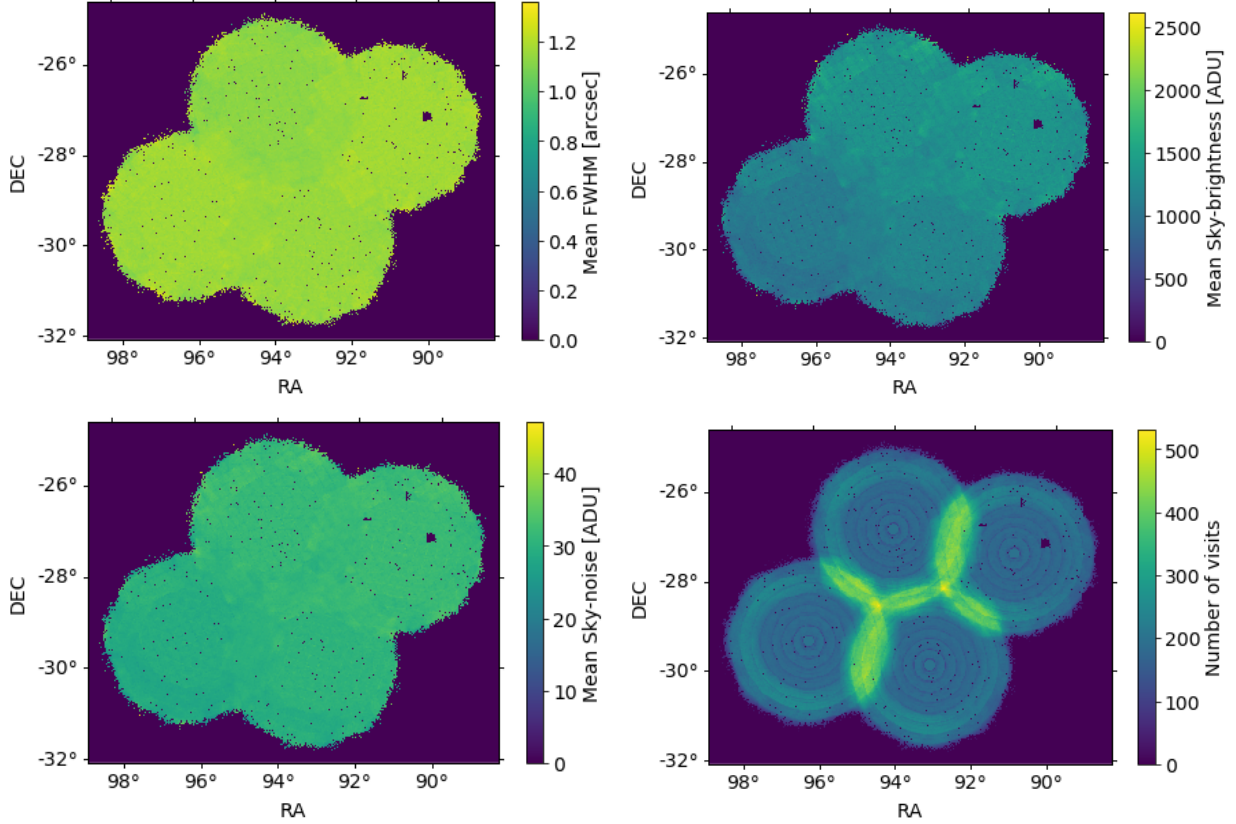


Figure 24. Same as Figure 23 for the undithered dataset. From top left to bottom right: Mean PSF FWHM, mean sky brightness, mean sky noise and number of visits. Note that we use the same extinction and stellar density maps changing the geometry of the mask.

the close neighborhood to bright objects our ability to detect faint sources diminishes. These faint sources are blended in the core or the tails of the brighter objects, resulting in a lower mean number of detected sources, as shown in Figure 20. The correction due to weighting by the fraction of area covered in each pixel has a considerable impact at low-scales, being larger than the statistical uncertainty in this regime. This showcases again the importance of considering the impact of blending in the low scale regime for LSST and should be carefully studied in future Data Challenges. The correction due to the presence of bright objects is comparable in both simulations.

9. CONCLUSIONS

End-to-end simulations are powerful tools for testing the overall performance of any current and future cosmological experiments like the LSST (Ivezic et al. 2008). They allow us to validate and improve on different parts involved in the data processing and analysis, as well as to model and improve our control of systematic uncertainties.

In this paper, we have presented a simulated (end-to-end) imaging dataset that resembles single-band, full-

depth LSST data, for the first data challenge in the LSST DESC. We simulated images using state of the art tools (*imSim*). We generated two different and complementary datasets, one with random dithers (*dithered*) and the other with no dithers (*undithered*). We processed these images with the LSST data management software stack and performed several quality assurance tests on its outputs. We checked that both the *dithered* and *undithered* are high-quality datasets and pass the LSST Science Requirements (Ivezić & the LSST Science Collaboration 2013) and the DESC Science Requirements (The LSST Dark Energy Science Collaboration et al. 2018) allowed by our design choices.

We studied different ways to relate the output catalogs to the inputs. In particular, we studied two different matching strategies. The first, using information about positions only and a second strategy involving positions and magnitudes. We showed that, for clustering analyses, adding information about magnitudes results in a lower incidence of spurious matches. The usage of matching strategies helped us define a clean sample suitable for clustering analyses. After cleaning the catalog, we found a small fraction ($\approx 3.6\%$) of artifacts, i.e.,

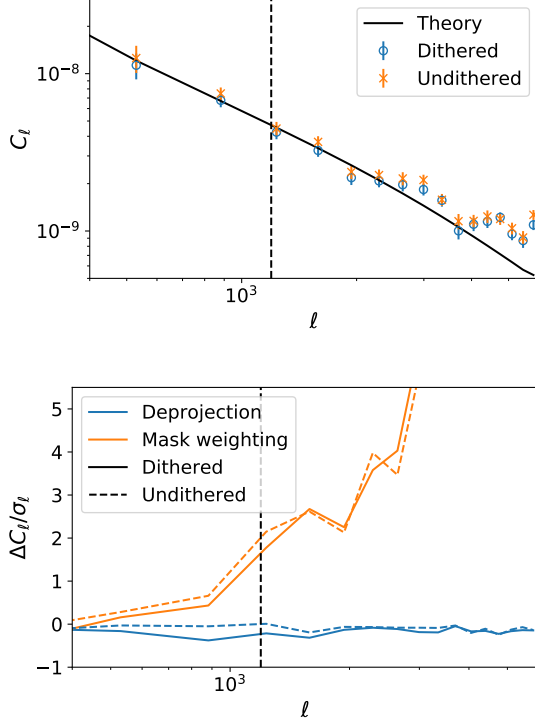


Figure 25. Top panel: Measured power spectra undithered (orange \times) and dithered (open blue circles) datasets with **NaMaster** corrected by systematics. The error bars are computed using jackknife. A theoretical prediction is shown as the solid black line to demonstrate the overall agreement between the measurements and the data. The vertical black dashed line corresponds to the minimum multipole considered for our analysis $\ell = 1/f_{sky}$. **Bottom panel:** Size of the correction in the power-spectra, ΔC_ℓ , relative to their uncertainty, σ_ℓ , due to deprojection (blue) of different observing conditions, and the correction due to weighting by the bright object mask (orange) for the dithered (solid lines) and undithered (broken lines) simulations. We see that the largest impact comes from the presence of bright objects, and that it is important to account for the area lost by masking via weighting.

objects with no counterpart in the input. We checked

the impact of these objects in the two-point clustering statistics, and demonstrated that our selection criteria is robust given the DC1 area.

However, careful selection criteria will be needed to enable accurate clustering analyses for the larger Y10 dataset. We anticipate that additional information coming from multi-band coverage and photometric redshifts, will help us to further refine the selection. We also realize that the matching techniques that we use in this study are likely insufficient for studies on blending or scales smaller than those considered in this work, since they do not include information about undetected sources present in blends. Therefore, an important research topic for these kinds of end-to-end simulations is to find efficient strategies to relate inputs and outputs.

Finally, we selected a high-completeness sample to perform clustering analysis in harmonic space and used **NaMaster** to deproject the impact of simulated foregrounds. The results of this analysis indicate that the simulated foregrounds have a low impact ($\leq 30\%$ of the statistical uncertainty) at the scales considered for our study ($1200 \leq \ell < 6000$) in both datasets. This is probably due to the simplicity of our foregrounds, and more complexity will be added in future data challenges. We also observe that the dithered simulation seems to be less affected by these foregrounds but, the difference is not statistically significant given the area of DC1. We also see that, in the aforementioned scale-range, the presence of bright objects has a larger impact on the power-spectra, $\approx 200 - 600\%$ of the statistical uncertainty, highlighting the impact of blending in LSST for small-scale analyses.

Finally, we have been able to perform an end-to-end test of our processing and analysis pipelines, which will allow us to better prepare to exploit future LSST data.

The methodology presented in this work will serve as the basis for future DESC data challenges, where we aim to perform multi-band studies in a larger area, analyze complementary image generation strategies (**PhoSim**), and increase the complexity of the foregrounds included.

APPENDIX

A. DETAILED ASTROMETRIC AND PHOTOMETRIC TESTS

We show the detailed astrometric repeatability in Figure 26, where detected stars that are detected at different visits are matched and their astrometric residual is shown as a function of their SNR.

ACKNOWLEDGMENTS

FJS thanks Amanda Pagul and David Alonso for useful comments and discussion. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We acknowledge the use of **Pandas**, **Dask**, **SciPy**, **Matplotlib**, **Jupyter**, **CCL**, **NaMaster**, **Healpy**, and **scikit-learn** as well as the LSST software stack. The work of FJS and DK was supported by the US Department of Energy award DE-SC0009920.

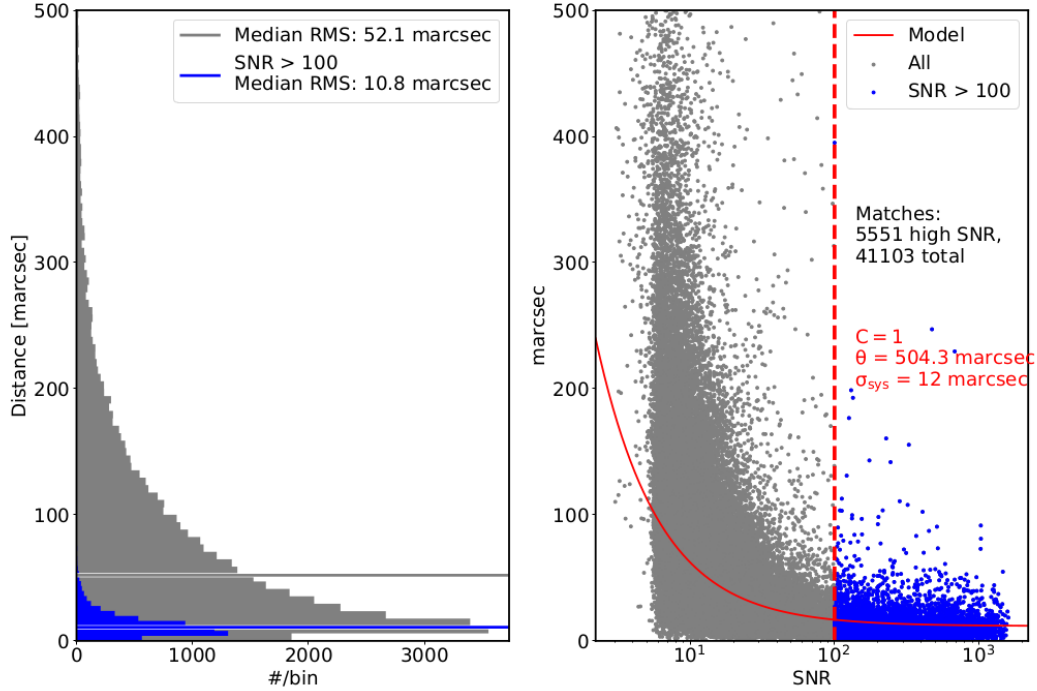


Figure 26. Detailed astrometric performance. (left) Histogram of the repeatability of distance between pairs of stars. (right) Astrometric variation as a function of SNR. We find a systematic floor of 12 milliarcsec.

The work of CWW was supported by the US Department of Energy High Energy Physics grant DE-SC0010007. The work of JC, RD, SD, TG, AJ, HK, PJM and BVK was supported by the U.S. Department of Energy under contract number DE-AC02-76SF00515. The work of RM was supported by the US Department of Energy Cosmic Frontier program, grant DE-SC0010118.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

Author contributions are listed below.

- J. Sánchez: led study
- C. W. Walter: led image generation
- A. Slosar: Participated in analysis and preliminary tests
- D. Kirkby: Participated and advised in analysis
- J. Chiang: One of the main developers of imSim, participated in image generation and data distribution
- T. Glanzman: Generated artificial images
- S. F. Daniel: Developed imSim and CatSim
- H. Awan: Participated in analysis and designed dithering strategy
- E. Gawiser: Participated in analysis and designed dithering strategy

W. M. Wood-Vasey: Ran validation software and wrote part of the document
 Y. AlSayyad: LSST Science Pipelines
 C. Burke: Contributions to PhoSim
 J. Cheng: Contributions to PhoSim
 S. Digel: Workflow and validation
 R. Dubois: Workflow and management
 M. Jarvis: Contributions to imSim and validation
 T. Johnson: Workflow, data generation and processing
 H. Kelly: DMstack librarian at NERSC
 S. Krughoff: Contributions to LSST Science Pipelines and DC1 workflow
 R. H. Lupton: LSST Science Pipelines, overall QA
 R. Mandelbaum: Provided feedback on DC1 design, analysis, and paper draft
 P. J. Marshall: Led the “Twinkles” DC1 pathfinder project.
 M. Mustafa: NERSC workflow
 E. -H. Peng: Contributions to PhoSim
 J. R. Peterson: PhoSim Development for DC1
 P. Price: LSST Science Pipelines
 G. Sembroski: PhoSim Development for DC1
 B. Van Klaveren: DC1 workflow
 M. P. Wiesner: Completed a study on astrometry in PhoSim making it possible to improve astrometry in PhoSim for DC1
 B. Xin: Contributions to ImSim and PhoSim

REFERENCES

- Alonso, D., Sanchez, J., & Slosar, A. 2019, MNRAS, 484, 4127, doi: [10.1093/mnras/stz093](https://doi.org/10.1093/mnras/stz093)
- Awan, H., Gawiser, E., Kurczynski, P., et al. 2016, ApJ, 829, 50, doi: [10.3847/0004-637X/829/1/50](https://doi.org/10.3847/0004-637X/829/1/50)
- Benn, C. R. 1983, The Observatory, 103, 150
- Bertin, E. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435
- Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, Publications of the Astronomical Society of Japan, 70, S5, doi: [10.1093/pasj/psx080](https://doi.org/10.1093/pasj/psx080)
- Bruderer, C., Chang, C., Refregier, A., et al. 2016, ApJ, 817, 25, doi: [10.3847/0004-637X/817/1/25](https://doi.org/10.3847/0004-637X/817/1/25)
- Brun, R., Hagelberg, R., Hansroul, M., & Lassalle, J. C. 1978, Simulation program for particle physics experiments, GEANT: user guide and reference manual (Geneva: CERN). <https://cds.cern.ch/record/118715>
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, doi: [10.1046/j.1365-8711.2003.06897.x](https://doi.org/10.1046/j.1365-8711.2003.06897.x)
- Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301, doi: [10.1086/587156](https://doi.org/10.1086/587156)
- Budavri, T., & Lored, T. J. 2015, Annual Review of Statistics and Its Application, 2, 113, doi: [10.1146/annurev-statistics-010814-020231](https://doi.org/10.1146/annurev-statistics-010814-020231)
- Chisari, N. E., Alonso, D., Krause, E., et al. 2019, ApJS, 242, 2, doi: [10.3847/1538-4365/ab1658](https://doi.org/10.3847/1538-4365/ab1658)
- Connolly, A. J., Peterson, J., Jernigan, J. G., et al. 2010, in Proc. SPIE, Vol. 7738, Modeling, Systems Engineering, and Project Management for Astronomy IV, 77381O
- Connolly, A. J., Angeli, G. Z., Chandrasekharan, S., et al. 2014, in Proc. SPIE, Vol. 9150, Modeling, Systems Engineering, and Project Management for Astronomy VI, 915014
- Coupon, J., Czakon, N., Bosch, J., et al. 2018, Publications of the Astronomical Society of Japan, 70, S7, doi: [10.1093/pasj/psx047](https://doi.org/10.1093/pasj/psx047)
- De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499, doi: [10.1111/j.1365-2966.2005.09879.x](https://doi.org/10.1111/j.1365-2966.2005.09879.x)
- de Ruiter, H. R., Willis, A. G., & Arp, H. C. 1977, A&AS, 28, 211
- Delgado, F., Saha, A., Chandrasekharan, S., et al. 2014, in Proc. SPIE, Vol. 9150, Modeling, Systems Engineering, and Project Management for Astronomy VI, 915015
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, ArXiv e-prints. <https://arxiv.org/abs/1611.00036>
- Elsner, F., Leistedt, B., & Peiris, H. V. 2017, MNRAS, 465, 1847, doi: [10.1093/mnras/stw2752](https://doi.org/10.1093/mnras/stw2752)

- Gonzalez, O. A., Rejkuba, M., Zoccali, M., Valenti, E., & Minniti, D. 2011, *A&A*, 534, A3, doi: [10.1051/0004-6361/201117601](https://doi.org/10.1051/0004-6361/201117601)
- Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, 464, 399, doi: [10.1051/0004-6361:20066170](https://doi.org/10.1051/0004-6361:20066170)
- Hirata, C., & Seljak, U. 2003, *MNRAS*, 343, 459, doi: [10.1046/j.1365-8711.2003.06683.x](https://doi.org/10.1046/j.1365-8711.2003.06683.x)
- Hirata, C. M., Mandelbaum, R., Seljak, U., et al. 2004, *MNRAS*, 353, 529, doi: [10.1111/j.1365-2966.2004.08090.x](https://doi.org/10.1111/j.1365-2966.2004.08090.x)
- Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, *ApJ*, 567, 2, doi: [10.1086/338126](https://doi.org/10.1086/338126)
- Ivezić, Ž., & the LSST Science Collaboration. 2013, LSST Science Requirements Document. <http://ls.st/LPM-17>
- Ivezic, Z., Tyson, J. A., et al. 2008, ArXiv e-prints. <https://arxiv.org/abs/0805.2366>
- Jarvis, M., Bernstein, G., & Jain, B. 2004, *MNRAS*, 352, 338, doi: [10.1111/j.1365-2966.2004.07926.x](https://doi.org/10.1111/j.1365-2966.2004.07926.x)
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, 673, 864, doi: [10.1086/523619](https://doi.org/10.1086/523619)
- Jurić, M., Kantor, J., Lim, K., et al. 2015, ArXiv e-prints. <https://arxiv.org/abs/1512.07914>
- Kriszianas, K., & Schaefer, B. E. 1991, *PASP*, 103, 1033, doi: [10.1086/132921](https://doi.org/10.1086/132921)
- LSST Dark Energy Science Collaboration. 2012, ArXiv e-prints. <https://arxiv.org/abs/1211.0310>
- LSST Science Collaboration. 2009, ArXiv e-prints. <https://arxiv.org/abs/0912.0201>
- Mandelbaum, R., Hirata, C. M., Seljak, U., et al. 2005, *MNRAS*, 361, 1287, doi: [10.1111/j.1365-2966.2005.09282.x](https://doi.org/10.1111/j.1365-2966.2005.09282.x)
- Mandelbaum, R., Miyatake, H., Hamana, T., et al. 2018, *PASJ*, 70, S25, doi: [10.1093/pasj/psx130](https://doi.org/10.1093/pasj/psx130)
- McKinney, W. 2010, in *Proceedings of the 9th Python in Science Conference*, ed. S. van der Walt & J. Millman, 51 – 56
- Miralda-Escude, J. 1991, *ApJ*, 380, 1, doi: [10.1086/170555](https://doi.org/10.1086/170555)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Peebles, P. J. E. 1973, *ApJ*, 185, 413, doi: [10.1086/152431](https://doi.org/10.1086/152431)
- Peterson, J. R., Jernigan, J. G., Kahn, S. M., et al. 2015, *ApJS*, 218, 14, doi: [10.1088/0067-0049/218/1/14](https://doi.org/10.1088/0067-0049/218/1/14)
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, 10, 121, doi: [10.1016/j.ascom.2015.02.002](https://doi.org/10.1016/j.ascom.2015.02.002)
- Sánchez, J., et al. in prep., Olber's Paradox Revisited – Effects of Overlapping Sources on Cosmic Shear Estimation: Statistical Sensitivity and Pixel-Noise Bias
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525, doi: [10.1086/305772](https://doi.org/10.1086/305772)
- Sérsic, J. L. 1963, *Boletín de la Asociacion Argentina de Astronomia La Plata Argentina*, 6, 41
- Sjöstrand, T., Mrenna, S., & Skands, P. 2006, *Journal of High Energy Physics*, 5, 026, doi: [10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026)
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629 EP
- The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv e-prints, arXiv:1809.01669. <https://arxiv.org/abs/1809.01669>
- Wolstencroft, R. D., Savage, A., Clowes, R. G., et al. 1986, *MNRAS*, 223, 279, doi: [10.1093/mnras/223.2.279](https://doi.org/10.1093/mnras/223.2.279)
- Yoachim, P., Coughlin, M., Angeli, G. Z., et al. 2016, in *Proc. SPIE*, Vol. 9910, *Observatory Operations: Strategies, Processes, and Systems VI*, 99101A