MNRAS **000**, 000–000 (0000) Preprint 8 May 2018 Compiled using MNRAS LATEX style file v3.0

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

J. Sánchez[1], C. W.Walter[2], H. Awan[3], E. Gawiser[3], A. Slosar[4], D. Kirkby[5], J. C. Chiang[6], T. Glanzman[6], S. F. Daniel[7], Y. AlSayyad[8], C. Burke[8], J. Cheng[8], S. Digel[6], R. Dubois[8], M. Jarvis[8], T. Johnson[8], H. Kelly[8], S. Krughoff[8], R. H. Lupton[8], R. Mandelbaum[8], P. J. Marshall[9], M. Mustafa[8], E. -H. Peng[8], J. Peterson[8], P. Price[8], G. Sembroski[8], B. Van Klaveren[8], M. Wiesener[8], B. Xin[8]

(LSST DESC Collaboration)

[1] *Department of Physics and Astronomy, University of California, Irvine, Frederick Reines Hall, Irvine, CA, U.S.A.*
[2] *Duke University, Department of Physics, Durham, NC, U.S.A.*
[3] *Department of Physics & Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Rd, Piscataway, NJ 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, USA*
[7] *Department of Astronomy, University of Washington, Seattle, WA, USA*
[8] *Affiliation*
[9] *Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA, USA*

8 May 2018

**ABSTRACT**

The success of the Large Synoptic Survey Telescope (LSST) as a dark energy experiment will depend on controlling systematic effects for the various cosmological probes. Simulations are critical for developing the methodology to estimate and mitigate these systematics. In the first Data Challenge from the LSST Dark Energy Science Collaboration, we evaluate the potential systematic effects that will affect observables, with an emphasis on galaxy clustering. We simulate LSST images using `imSim`. Simulated images are then processed, combined, and analyzed using the current version of the LSST Data Management pipeline. Here we characterize the resulting systematics and implement corrections. Our results demonstrate that we can generate realistic LSST-like simulated images and control the systematic effects, after processing these images, at a sufficient level to enable major advances in our knowledge of dark energy and cosmology. The methodology presented here can be easily translated to current and future imaging surveys.

**Key words:** LSS , Data challenge, Systematics

## 1 INTRODUCTION

The increase in statistical power from recent cosmological experiments makes the modeling, and mitigation of systematic uncertainties key to extracting the maximum performance and producing competitive analyses. More traditional in high energy particle physics (Brun et al. 1978), (Sjöstrand et al. 2006), end-to-end simulations provide a unique framework to model systematics and streamline processing and analysis pipelines given that we have complete information about the inputs and outputs. With the larger availability of computational resources this approach has also been extended to photometric redshift galaxy surveys (Suchyta et al. 2016; Bruderer et al. 2016), and a similar effort is undergoing in spectroscopic surveys such as DESI (DESI Collaboration et al. 2016). For surveys like the LSST (Ivezić et al. 2008), where the expected data volume is very large, and where a highly stringent control of the systematic uncertainties is required, producing these kind of end-to-end simulations becomes necessary. With $\sim 50$ PB of raw data and $\sim 40$ billion objects (Ivezić et al. 2008) the data handling by itself becomes challenging. There are many approaches to generate these end-to-end simulations. Usually, one would start from a source catalog with sources parametrized somehow (*e.g.* sizes, shapes, fluxes), render images from these sources, adding some observational effects and, finally, use

© 0000 The Authors

some software to detect and measure different properties on these images. Depending on the approach used to generate these images, some difficulties may arise when trying to relate inputs and outputs, especially, in a situation like LSST with its large data volume.

In this paper, we use simulated images that resemble the data that will be produced by LSST (Ivezić et al. 2008) after 10 years of operation in $r$-band using state of the art tools. We characterize the products of this process, measuring photometric, astrometric and clustering properties of the sample. These products encompass single-visit and coadded calibrated exposures (i.e., flattened, background subtracted, etc) and source catalogs that add up to $\sim 225TB$. They are the result of two different simulations that will be introduced later.

This paper is structured as follows: In Section 2 we describe the inputs for our simulated images. In Section 3 we discuss the dithering strategies used for this study. In Section 4 we describe the process to generate our LSST-like artificial images. In Section 5 we describe the data products generated after processing and performs several validation tests. In Section 6 we present the methodology followed to perform the clustering analyses on the simulated data products. Finally, in Section 7 we present some concluding remarks.

## 2 IMAGE GENERATION: INPUT CATALOG

Please, Daniel and/or Simon check this 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 definitely not be realistic, and we will not able to capture some correlations present in real data. On the other hand, using a N-body simulation as the input to generate artificial images allows us to study all the aforementioned effects. This is why we used the `CatSim` Reference? catalog as our input. This catalog uses the dark matter haloes from the Millennium simulation (Springel et al. 2005) and a semi-analytic baryon model described in De Lucia et al. (2006). The semi-analytic model features radiative cooling, star formation, the dynamics of black holes, supernovae, and AGNs and was adjusted to mimic the luminosity, color, and morphology distributions of low redshift galaxies. `CatSim` was generated by constructing a lightcone, covering redshifts $0 < z < 6$ from 58 500 $h^{?1}$ Mpc simulation snapshots. The final catalog comprises a 4.5×4.5 degree footprint on the sky (sufficient to cover a single LSST field-of-view) and samples halo masses over the range $2.5 \times 10^9$ to $10^{12}$ $M_\odot$.I copied this from https://www.lsst.org/scientists/simulations/catsim, is that fine?

For DC1, this catalog was tiled to generate a $\sim 40$ deg$^2$ footprint. This approach introduces some periodicity that induces some spurious correlations in our sample, however, we confirmed that these effects appear at scales larger than we are able to map given the DC1 area. For all sources, a spectral energy distribution (SED), is fit to the galaxy colors

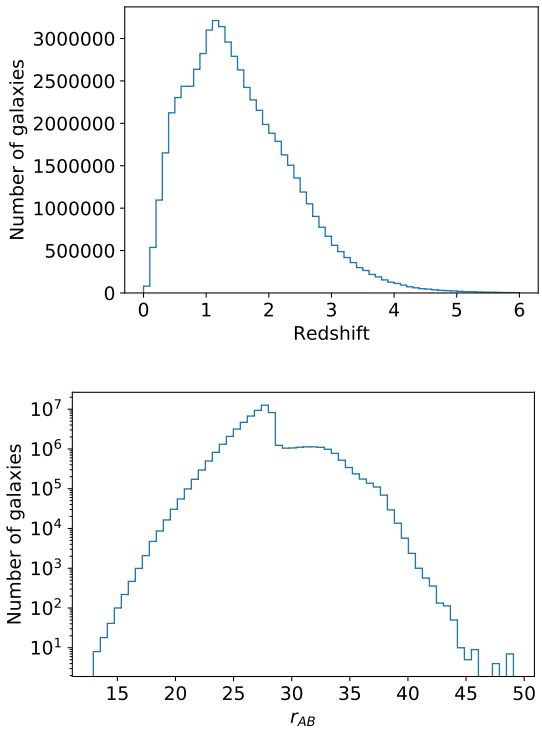


**Figure 1.** Redshift (top) and magnitude (bottom) distribution for the galaxies used as inputs for DC1.

using Bruzual & Charlot (2003) spectral synthesis models. The De Lucia et al. (2006) catalog includes BVRIK magnitudes and dust values for the disk and bulge components of each galaxy as well as radii, redshift, coordinates, stellar age, masses and metallicities. 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). Bulge-to-disk ratios and disk scale lengths are taken from (De Lucia et al. 2006). Half-light radii for the bulge components are derived from the absolute-magnitude vs half-light radius relation given by Gonzalez et al. (2011). Colors and stellar mass of the AGN host galaxies are estimated from the AGN luminosities.

Stars are represented as point sources and are drawn from the Galfast model (Jurić et al. 2008). Galfast generates stars according to density laws derived from fitting SDSS data to a model of a thick and thin disk, and a halo. Each star is assigned a metallicity, proper motion, and parallax.

In total, the input catalog contains approximately 63.1 million sources of which, 61.3 million are galaxies whose redshift and magnitude distributions are depicted in Figure 1 and 1.8 million are stars. Using this catalog we generate images covering an area of approximately 40 square degrees (4 pointings) in r-band to LSST full depth (10 years). The final footprint can be seen in Figure 2

## 3 DITHERING STRATEGY

Please, Humna and/or Eric check this section.

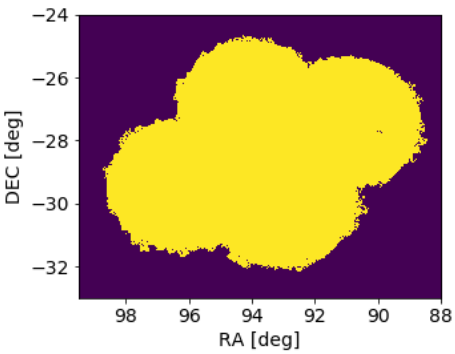


**Figure 2.** Footprint of the DC1 dataset. We simulate 4 LSST full focal plane pointings which roughly corresponds to 40 square degrees.

Different dithering strategies impact survey uniformity and systematic effects in photometric redshift surveys (Awan et al. 2016). Simulating two different strategies is a beneficial check to determine the most effective strategy to mitigate systematic effects and yield better cosmological constraints. For this data challenge we use two different dithering strategies: the *undithered* which points the telescope at the same position and orientation for all exposures in a given field; and the *dithered* which consists in small random rotational and translational dithers for each pointing/exposure. This dithering strategy is stored in the OpSim[1] database presented in ref. (Awan et al. 2016). However, with dithering, we might have some visits where part of the focal plane lies outside of our region of interest. We also simulated these visits, however, in order to save computational resources, we decided not to simulate the chips that lie outside of the region of interest.

## 4 IMAGE GENERATION: PIPELINE

The artificial generation of astronomical images is a complex and computationally demanding process. In the recent years there is a big effort in the community in order to create software that allows more realistic and fast image generation like `BALROG` (Suchyta et al. 2016) or `UFIG` (Bruderer et al. 2016). In our case, we model the input sources using imSim which uses GalSim (Rowe et al. 2015) as a library driven by a LSST specific application. There are other approaches consisting in running a full photon-shooting simulation using PhoSim (Peterson et al. 2015). This approach will also be used in future Data Challenges. The former has a speed advantage but the latter fully traces each photon coming from the sources through the atmosphere and the instrument, increasing the level of realism with a fairly high computational cost. These two approaches can potentially allow us to focus on different systematic effects and science cases. -¿ Rewrite

[1] `https://www.lsst.org/scientists/simulations/opsim`

### 4.1 imSim

imSim follows the path of the fast image simulations in refs. (Suchyta et al. 2016; Bruderer et al. 2016). imSim uses GalSim (Rowe et al. 2015) as a library to generate images that resemble LSST individual visits. Instead of the planned two 15 seconds exposures planned (Ivezić et al. 2008) we generate a single 30 seconds image to simplify the data handling. We simulate each CCD of the focal plane individually but there are no instrumental effects nor variability in the optical model across the focal plane except for the sky. We used an ESO sky model and the sky brightness values come from OpSim reference?[2]. The PSF model that we use is a Gaussian for the system with a FWHM that depends on the airmass [3] plus a Kolmogorov profile to model the atmosphere that is airmass dependent as well [4]. The airmass model is taken from (Krisciunas & Schaefer 1991);

$$X = (1 - 0.96 \sin Z)^{-0.5}, \tag{1}$$

where $X$ is the airmass, and $Z$ the angular distance to the zenith.

In this simulation we generated three different types of objects: Stars which are modeled as PSF-like objects; galaxies, which are modeled as Sérsic profiles (Sérsic 1963) (bulge plus disk) using the parameters given by CatSim; and AGNs that are also modeled as very narrow Gaussian profiles. In this version of the catalog we decided to not include variability in AGNs to simplify the analysis. The zeropoints are computed using the model presented at (Ivezić et al. 2008) and the throughputs at `https://github.com/lsst/throughputs` David and I used the throughputs from the repo to compute these zeropoints. Should I add a table with them?. We clip the objects at magnitude 10 to simulate saturation.

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 typical time to simulate each CCD is secondsCheck how long it took to do one chip?/Total imSim time.

### 4.2 Image processing pipeline

The outputs of these simulations are then processed using the LSST data management (DM) stack (Ivezić et al. 2008; LSST Science Collaboration 2009; LSST Dark Energy Science Collaboration 2012; Bosch et al. 2017; Jurić et al. 2015) . This is an open source high-performance data processing and analysis system intended for use in O/IR survey data. The code can be found at `dm.lsst.org` and `pipelines.lsst.io`. The raw, uncalibrated single exposures are used as inputs. The software performs the reduction, detection, deblending and measurement on individual visits and coadds producing the level 2 data products (Jurić et al. 2015). Although it is still under development, most of the pieces already meet the science requirements. The DM stack provides calibrated images and source catalogs for the individual visits, and coadds stored in `FITS` files. In total,

[2] We used the values at OpSim's run `minion_1016`
[3] From LSST-20160 eqn. (4.1)
[4] From LSST-20160 eqn. (4.2)

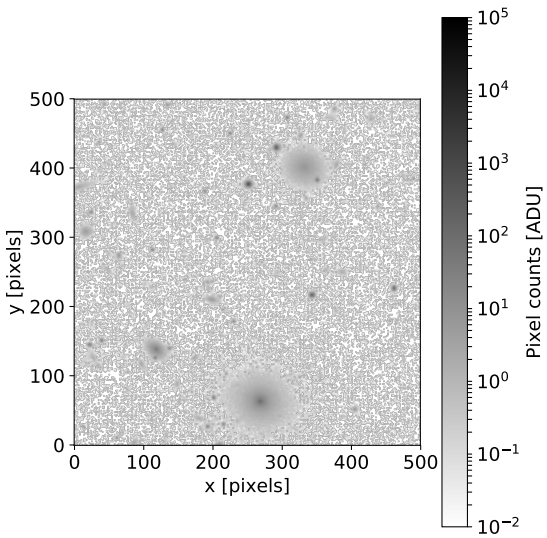


**Figure 3.** Example of a 500 × 500 pixel cutout from a full depth coadd. We can appreciate the large density of simulated objects.

we detect and measure ∼ 10.6 million objects with position, flux and shape information. We activated optional extensions for the pipeline to include `CMODEL` fluxes (see Bosch et al. (2017) for more details) and `HSM` shapes (Hirata & Seljak 2003; Mandelbaum et al. 2005).

An example of a coadd image can be seen at Figure 3.

## 5 OUTPUT CATALOGS AND VALIDATION

After being processed, the catalogs are accessible by DESC collaborators and stored at the National Energy Research Scientific Computing Center (NERSC)[5]. We generate `pandas dataframes` (Mckinney 2011) and databases for each one of the coadds and input catalogs in order to be accessed by the collaborators and perform their own analyses. The output catalogs contain 10.6 million objects covering an area of ∼ 43 deg$^2$. The catalogs include information about position, size, shape and magnitude for every object. They include several flags that give information about the presence of interpolated/saturated pixels in the objects or whether these objects are close to the edge of a CCD or not.

In order to check the level of realism and the accuracy of the processed catalogs we perform several quality assurance tests.

### 5.1 Astrometry checks

Biases in astrometry can potentially affect both clustering and weak lensing measurements. These biases can have different origins: PSF mis-characterization, non corrected sensor effects, incorrect modeling of proper motion for the measured objects and presence of blended sources are among the most common sources.

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

#### 5.1.1 External checks

As we have already mentioned, one of the big advantages of using simulations is that we have access to the *true* underlying information. We will use this information to check the precision of the astrometric measurements in single exposures and co-adds. For these studies we will select stellar objects. In order to do so, we use the classifier included in the LSST software stack[6] and choose objects with `base_ClassificationExtendedness_value==0`. We also require that `deblend_nChild==0` to ensure that the objects are completely deblended, i.e., they are primary sources. We match these objects to the stellar sources in the input catalog. In both cases we will use a `KDTree` (Pedregosa et al. 2011) to retrieve those objects in the input catalog that are in a radius of 0.2 arc-seconds (one pixel) of those detected in the output catalog and select the match that is closest in magnitude. We only consider sources which have a magnitude difference smaller than 0.02 magnitudes. We will discuss matching strategies in more detail at Section 6.1.

We selected a representative single visit (visit number 270675 for the imSim dithered run) and calculated the difference between the measured and the input positions. These are represented in Figure 4. We can see that both RA and Dec distributions are compatible with each other, meaning that there are no anisotropies in the detection, as expected from the inputs. However, we find that the distributions are asymmetric and that the median is not zero. This effect is even more noticeable when we accumulate visits as in Figure 5, where we accumulated the results for 50 randomly selected visits of the imSim dithered run. This effect is also present in the undithered imSim run. We also checked the dependence the mean astrometric residual with the magnitude of the objects as shown in Figure 5 where a mild bias for the brightest objects can be seen. This bias is smaller than 15 mas, much smaller than the resolution of the input N-body simulation. This means that the two point clustering statistics will not be affected by this bias. This bias arises from the fact that the objects in the images include proper motion, which is unaccounted for in our *truth tables*.

We also wanted to check if there is a preferred orientation for the differences between the input and output position in a single visit. Using the same visit as before we show the astrometric residuals in Figure 6. In this Figure we can see that the astrometric residuals do not show any noticeable structure and appear to be mostly random with the largest contributions close to the CCD edges.

#### 5.1.2 Internal checks

Another important test is to ensure the internal consistency of the astrometric solutions between different exposures. To do that we selected a small region of the co-added area (*patch*) and compared the positions of the objects detected

[5] http://www.nersc.gov

[6] To see more details about the classifier refer to section 4.9.10 at (Bosch et al. 2017)

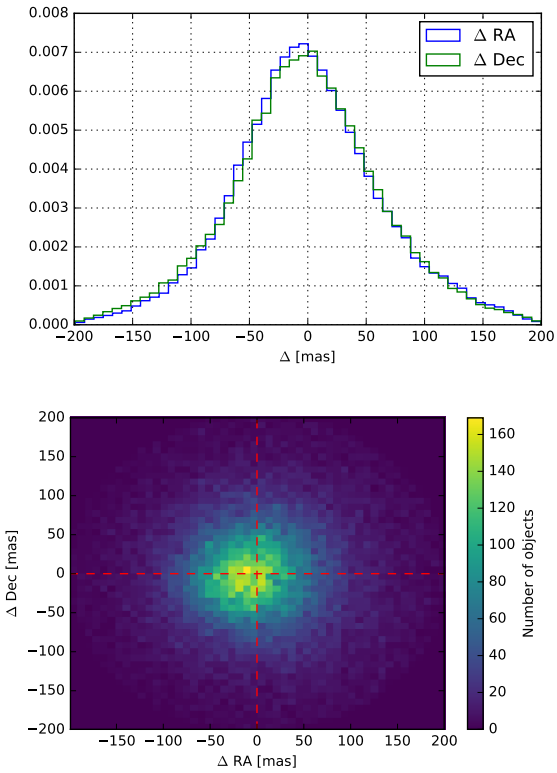


**Figure 4.** Left: Distribution of the difference $\Delta = X_{measured} - X_{input}$ in RA (blue) and Dec (green) coordinates. We cannot appreciate any differences between these, however we see that there is median is not at zero $\Delta_{median} \approx -2$ mas. The histograms are normalized such that the total sum of the counts is equal to one. Right: 2D histogram showing the bivariate distribution of the difference in RA (horizontal axis) and Dec (vertical axis). We selected one random representative exposure (visit number 270675 for the imSim dithered run). The effect is similar for the undithered imSim run and for the PhoSim run.

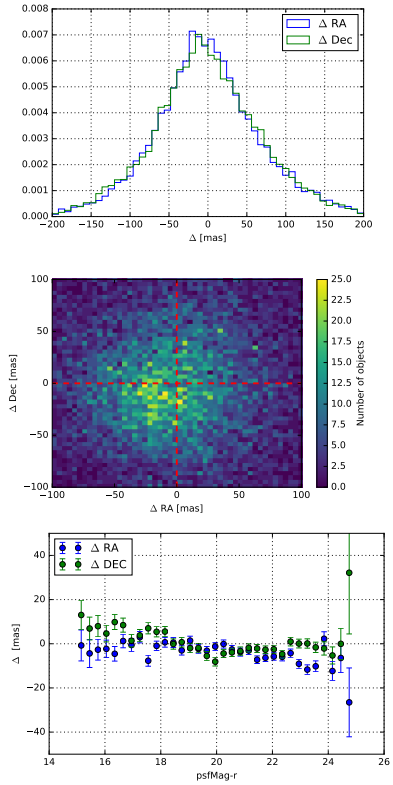


**Figure 5.** Left: Distribution of the difference $\Delta = X_{measured} - X_{input}$ in RA (blue) and Dec (green) coordinates as in Figure 4 but accumulating the results for 50 randomly selected visits from the imSim dithered run. Middle: 2D histogram showing the bivariate distribution of the difference in RA (horizontal axis) and Dec (vertical axis). Right: Mean astrometric residual as a function of magnitude for RA (blue) and Dec (green). These distributions are similar for the undithered imSim run and for the PhoSim run.

in the co-add image with the positions of objects detected in individual exposures that overlap with that patch.

In particular, we randomly chose 10 catalogs from individual exposures and looked for objects that fulfilled the following criteria:

- `deblend_nChild==0`, this means that the object has been completely deblended (it is a primary match).
- `base_PixelFlags_flag_edge==0`, which means that the object is not close to an edge.
- `base_PixelFlags_flag_interpolatedCenter==0`, the object does not have any interpolated pixels in its center.

Note that in this case we are not requiring the objects to be classified as stars (we are omitting the cut in `base_ClassificationExtendedness_value`) but we are adding some cuts to ensure that the objects were properly measured. Once we perform our selection, the next step is to match the objects in the different exposures. To do so we use the matching algorithm included in the LSST software stackReference? and calculate the mean of the difference between the position of each source in the co-add, $X_{coadd,i}$ and

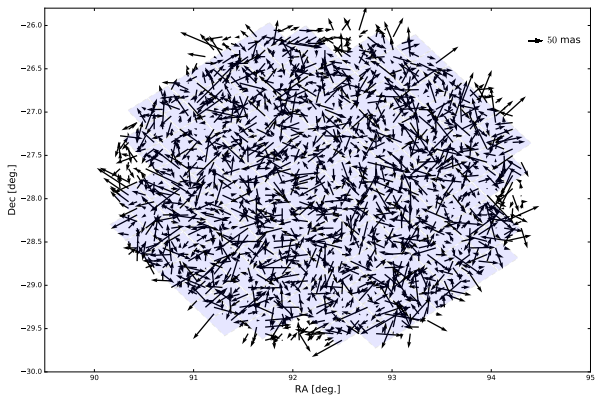


**Figure 6.** Astrometic residuals measured in visit 270675 from the imSim dithered run. The light blue squares represent the CCD chips in the LSST focal plane. The base of the arrow is on the input matched object. The arrows have been re-scaled 10,000 for visualization purposes.

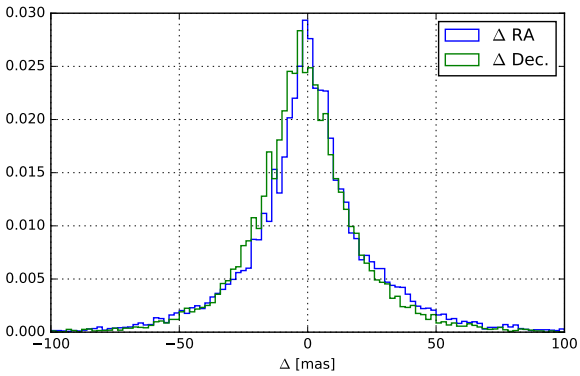


**Figure 7.** Distribution of the mean difference in position (RA:blue, Dec:green) between the coadd and the different individual exposures where each source has been detected.

the position of the matched object in each of the exposures where it has been detected, $X_{visit_j,i}$ for $j \in [1, 10]$, i.e,

$$\Delta = \langle X_{coadd,i} - X_{visit_j,i} \rangle \quad (2)$$

we only consider sources that have been detected in at least 5 exposures. The resulting distribution is shown in Figure 7 where we see that is noticeably narrower than those shown in Figure 4 and Figure 5 and no apparent bias is found showing that the processing is consistent.

## 5.2 Photometry checks

As we did in previous sections we perform two different tests to assess the quality of our simulations: first we compare our output catalogs to the inputs, and second we check the consistency between different visits for the same objects.

### 5.2.1 External checks

We want to analyze the accuracy of the magnitude measurement comparing the input catalog with the output catalog. This process not only checks the accuracy of the measurement pipeline, but it also tests the quality and level of realism of the image generation pipeline. For these external checks we use again the same 50 randomly selected visits as we did in the previous section to check the astrometric residuals. To study the photometric residuals we change slightly the matching strategy from previous sections. In this case we eliminate the threshold in magnitude difference so, we just look for the input source that is closest in magnitude in a 0.2 arc-seconds radius around each detected source. In Figure 8 we can see the distribution of the photometric residuals. We see that this distribution gets wider as we go fainter (as expected) and that the 0.02 magnitude selection cut was a good proxy to ensure that we account for most sources properly matched.

### 5.2.2 Internal checks

We also checked the consistency of the photometry between different exposures. Using the same approach and sample

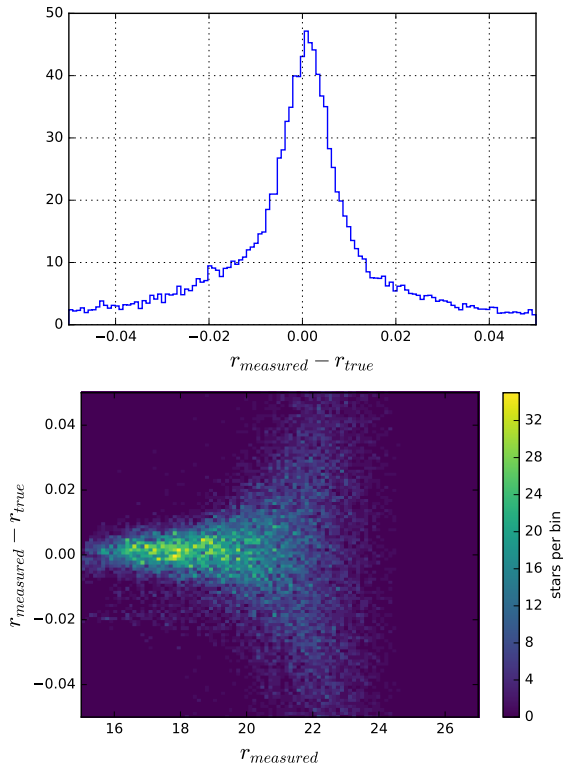


**Figure 8.** Left: Distribution of magnitude difference between the input and output catalogs. Right: Difference magnitude between input and output catalogs as a function of the measured magnitude. We considered 50 visits randomly selected from the imSim dithered run. We find similar results for the imSim undithered run and for the PhoSim run.

presented in Section 5.1.2 we compared the mean difference between the coadded source and 10 random single-visit images. In principle, we would expect that some objects will show differences between different epochs due to their intrinsic variability, however, these effects are not found in the inputs of the simulation for any of the objects, thus, the differences between different epochs will come mainly from statistical fluctuations and different observing conditions. The results from these checks can be seen in Figure 9.

In this figure we can see that the mean of the distribution is compatible with zero and that most sources have a magnitude difference lower than 25 mmag given the narrow peak. The presence of long tails are likely due to blends and mismatching or artifacts.

## 5.3 PSF checks

In order to ensure the accuracy of shape measurements a robust and accurate estimation of the PSF is required. In the case of imSim, we have a perfectly known input PSF that only depends on the airmass at the time of observation. We can check if the measurement pipeline can reconstruct the PSF and at which level of precision. To do that, we select 200 visits randomly, retrieve the measured PSF using the pipeline, and compare this to the input model given the observing conditions and compute the residual. We obtain

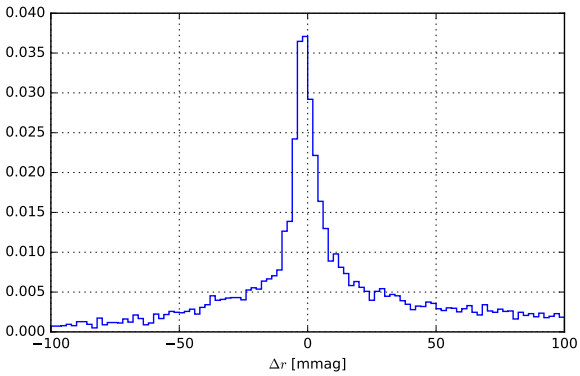


**Figure 9.** Distribution of the mean difference in magnitude between the coadd and the different individual exposures where each source has been detected. The mean of the distribution is consistent with zero.

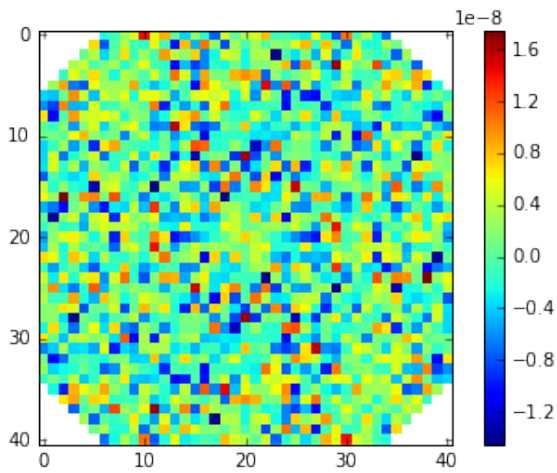


**Figure 10.** Average difference between the input PSF model (41×41 pixels, normalized to unity) and the measured PSF (with the same normalization) for 200 visits.

the residual depicted in Figure 10, where we can see that the PSF is recovered correctly with a high level of precision and accuracy.The very small residual in the center and the tails are dominated by noise.

## 6 ANALYSIS OF THE SIMULATED DATASET

In this section we are going to analyze the clustering properties of our output catalog from the end-to-end pipeline and compare to the input catalog.

### 6.1 Matching input and output catalogs

Potentially, in end-to-end simulations, one can trace back each measured photon to its corresponding source in order to characterize the measurement process. However, this is not always possible for several reasons. For example, a simple solution could be storing each postage stamp for the different sources in different exposures. The problem with this approach is that it creates a large overhead in an already big dataset. Generating efficient strategies to perform this bookkeeping and connecting input and output catalogs is then, an important research topic for these kinds of end-to-end simulations.

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 and performs reasonably well when blending is low, since the probability of confusing two objects is small. However, when blending is high, this approach might be not enough and matching other quantities like flux or shape can become useful, even though these can introduce some subtleties, an example being, how to weigh positions versus fluxes (and/or shapes) and whether or not we should consider the correlations between the two.

We compare two different matching strategies: simple spatial matching, where we match the find the objects in the truth catalog closest to the detected objects; and spatial matching including magnitude matching, where for each detected object we find which objects from the input catalog lie within one pixel radius ($0.2''$). 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 arbitrarily chose 0.5. If any of these conditions are not fulfilled, then the object is thrown away.

In the former approach all detected objects are matched to an object in the input catalog and, in principle, the two point statistics should be preserved between these two catalogs. However, in the latter approach, we can potentially bias the results since we are throwing away galaxies where the accuracy in the magnitude measurement is low, regardless of the precision. To compare both approaches we check the possible biases in the measured magnitude $\Delta\mathrm{mag} = \mathrm{mag}_{meas} - \mathrm{mag}_{true}$. In Figure 11 we can see that spatial matching is enough to get a good match in magnitude as evidenced by the median value of $\Delta$mag compared with the spatial+magnitude matching. This figure also shows that spatial matching is more affected by the presence of outliers than the spatial+magnitude matching, as expected by the clipping performed on the latter. The evolution of the mean and median $\Delta$mag with the magnitude evidences that brighter objects with magnitudes $\sim 18-21$ appear usually brighter and, on the other hand, fainter objects with magnitudes $\sim 23-25$ appear dimmer, this is likely a consequence of the deblender assigning a larger flux to the brightest members of the blend, while the fainter members seems to be losing flux. Finally, since for the faintest end (mag $\sim 26$) we can only detect those that appear brighter than they actually are and, thus, there is significant negative value for the median and mean $\Delta$mag.

We also compared the power-spectra of the detected objects and the matched input objects in Figure 12. In this figure we can see that the spatial matching recovers really well the power-spectrum, with some wiggling at very low scales due to mismatches and small differences in position between the detected and the input catalogs. In the case of spatial+magnitude matching, we see that the power for the matched catalog is consistently lower. This is due to the galaxies that are detected but not matched to any object in the true catalog, since this effect disappears when we just consider the objects that have been matched.

The problem with these matching techniques is the lack

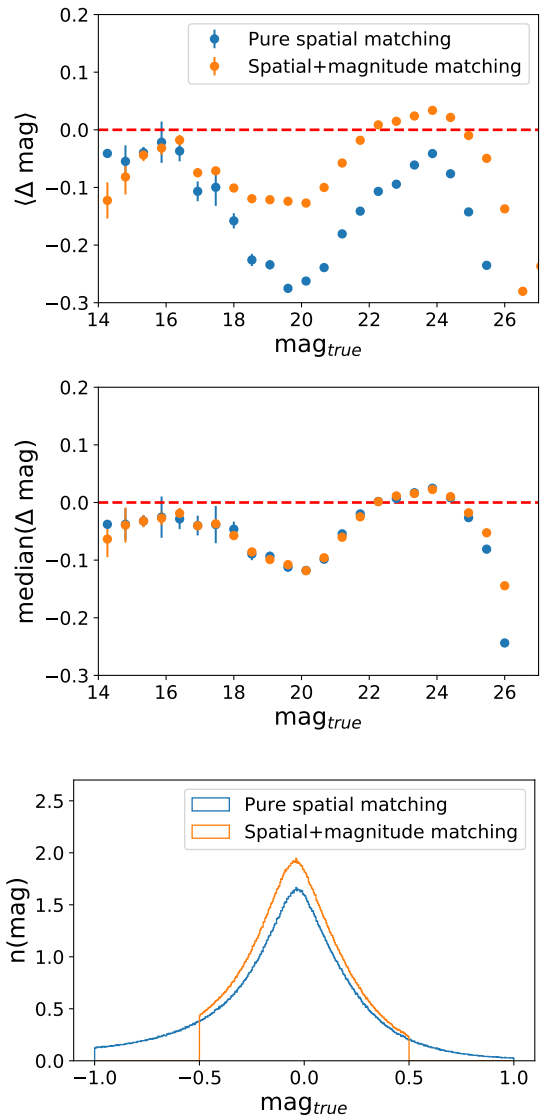


**Figure 11.** Mean (top), and median (middle) of Δmag as a function of the true magnitude, with pure spatial matching (blue) and spatial+magnitude matching (orange). Bottom: Histogram of Δ mag.

of very small scale information, and details about blended objects and groups, therefore, for analyses involving blending and/or very small scale information, different approaches should be considered.

### 6.2 Generating depth maps

In order to estimate the depth in the coadd catalogs we select the stars by using `base_ClassificationExtendedness_value==0`. This ensures selecting PSF-like objects. After this, we have two different approaches:

(i) In the first approach we generate a HEALPix[7] (Górski et al. 2005) map containing the PSF-like objects detected with a signal-to-noise ratio (SNR) higher or equal than 5 (`base_PsfFlux_flux/base_PsfFlux_fluxSigma>=5`), and we assign to each pixel the value of the dimmest object contained in it.

(ii) The second procedure also generates a HEALPix (Górski et al. 2005) map containing the PSF-like objects. Then in each pixel, we compute the median SNR as a function of the magnitude and get the magnitude at which SNR is the closest to 5.

These two procedures yield very similar results (within $\sim 4\%$) as it can be seen in Figure 13. We checked that the maps built selecting galaxies instead of stars are compatible as well. We will select our footprint according to the depth map using the second methodology which can be seen in Figure 14.

[7] http://healpix.sf.net

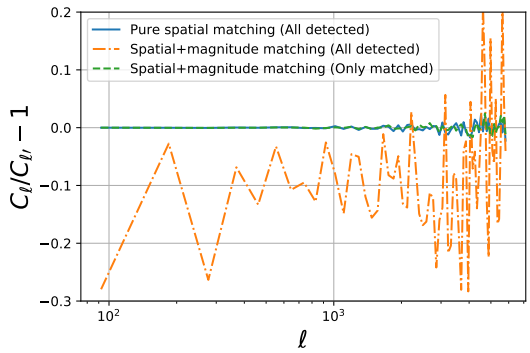


**Figure 12.** Relative difference between the power-spectrum measured for the detected objects and the matched objects in the input catalog with spatial matching compared to all detected objects (blue, solid line), spatial+magnitude matching compared to all detected objects (orange, dash dotted line), and spatial+magnitude matching compared to matched objects using this schema (green dashed line).

### 6.3 Bright objects masking and data selection

Bright objects produce significant effects in the image that affects the detection and measurement of neighboring objects. Some examples of the effect of these objects are: saturation of the CCD, large diffraction spikes, obscuration of neighboring sources, etc. Thus, masking a region around these sources makes a more complicated footprint but greatly simplifies the analysis of systematic effects. In order to evaluate the effect of this sources, we use the stars from the input catalog, we divide them in different magnitude bins, and we count the detected objects in a given radius. The main problem with this methodology is that the points are correlated, the number of stars in an aperture of radius $\theta_1 > \theta_0$, depends on the number of stars on the aperture with radius $\theta_0$. Therefore, it is better to use a differential measurement and this is what we do in Figure 15. We can see that there is an excess of sources at distances lower than 10 arcseconds for the brightest stars. This is mainly due to the existence of fake sources around these bright objects. The brighter sources also have a larger noise, the DM deblender models the brightest source and

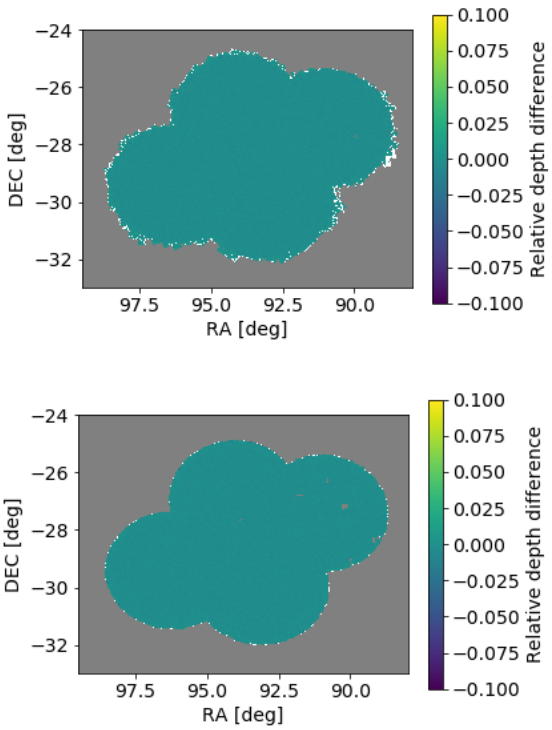


**Figure 13.** Relative difference between the depth calculated using the two methods presented in the text for the dithered (top) and undithered (bottom) fields.

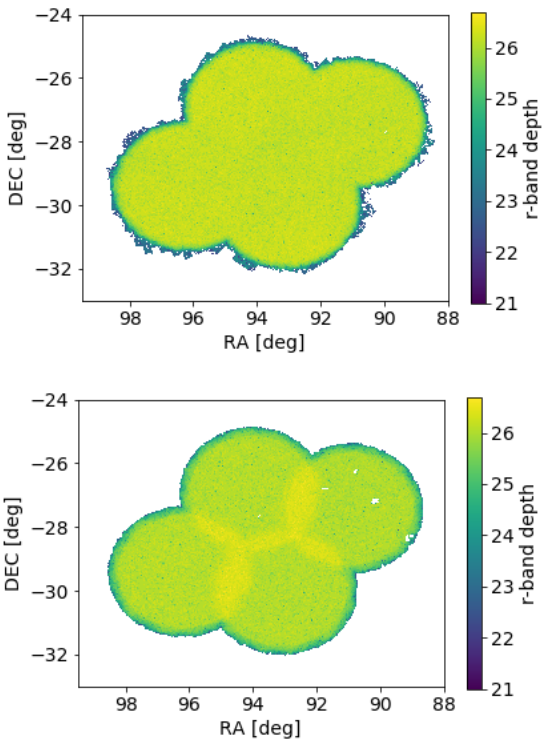


**Figure 14.** 5-$\sigma$ depth for the dithered (top) and undithered (bottom) fields.

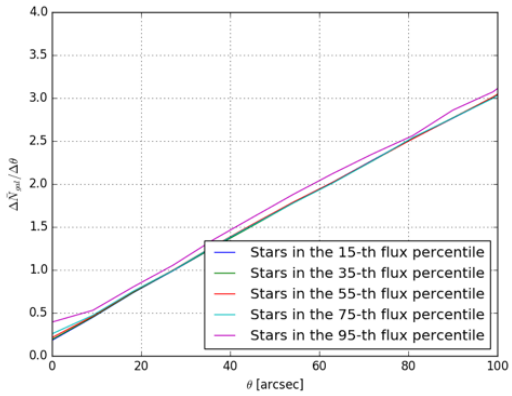


**Figure 15.** Mean increment in the number of detected objects around stars in the 95th flux percentile (magenta), the 75th percentile (cyan), the 55th percentile (red), the 35th percentile (green) and the 15th percentile (blue). We see that for the brightest percentiles there is no obscuration present but a overdensity of targets. These are mainly noise peaks identified as point sources.

subtracts this model looking for fainter sources that can be blended together with it. However, sometimes there are large noise peaks that have not been subtracted. These large noise peaks are detected in later stages as individual (point-like) sources giving as a result these fake sources. However, we do not see any obscuration present at the scales that we are interested (1 arcmin and higher) so it seems that no masking is required but we need to select the objects for our analysis carefully.

We checked the detection efficiency of galaxies in our sample. In order to do so we selected galaxies in the catalog using the variable `base_ClassificationExtendedness_value` as a star/galaxy separator as in ref. (Bosch et al. 2017). Objects where this variable is 1 are more likely to be galaxies, whereas the objects where it is 0 are more likely to be stars. We also made some quality cuts by selecting the objects with `detect_isPrimary==True`. This ensures that the object has been fully deblended and that the detection was not close to the edge of a coadded image. The results can be seen in Figure 16 where we computed the ratio of detected objects classified as galaxies and the input number of galaxies, $\varepsilon$ as a function of magnitude. We checked this ratio for both the dithered and undithered catalogs and in pixels where the depths were higher than 25.7 and 26.25. We used `CMODEL_MAG` as the reference magnitude for the detected objects and the true magnitude for the input galaxies.

Looking at Figure 16 it seems that selecting galaxies with `base_ClassificationExtendedness_value==1`, `detect_isPrimary==True` and `CMODEL_MAG` $>$ `21` is enough to get rid of most potential problems with fake detections. We confirmed this by performing visual inspection on the images. We noticed that fake sources happen close to very bright objects. On top of that, it seems that selecting `CMODEL_MAG` $<$ `25.3` ensures good level of completeness ($\sim$ 80%). Uniformity is ensured by selecting the galaxies that lie within HEALPixels where the depth is greater or equal than 25.3. After these selection cuts we end up with

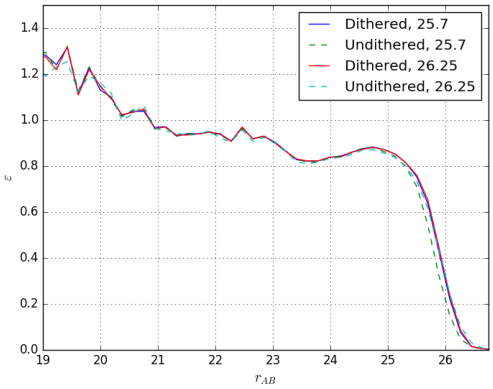


**Figure 16.** Number of detected objects classified as galaxies divided by the number input galaxies as a function of magnitude. We see that the detection efficiency is larger than 80% for the most part and a considerable fraction of objects with magnitude lower than 20 are being misclassified.

4.4 million galaxies for the dithered field and 4.3 for the undithered field.

### 6.4 Clustering results

In this section we analyze the two point clustering statistics for both the dithered and undithered catalogs in real and harmonic space and check the consistency between the input and measured observables. When comparing the input and output catalogs some subtleties arise. Let us think of a galaxy that has been measured with magnitude $m_{meas}$ and with true magnitude $m_{true}$ and we choose a magnitude threshold $m_{threshold}$ to define the sample of interest. If the luminosity function were flat, and assuming that the errors in magnitude are symmetric (which is not true) we would have as many objects with $m_{meas} < m_{threshold}$, but $m_{true} > m_{threshold}$ as objects with $m_{meas} > m_{threshold}$ but $m_{true} < m_{threshold}$, and this wouldn't change very much the clustering properties of our sample and we could just compare directly the samples using the magnitude threshold in true magnitude and measured magnitude. However, this is not the case, given that the luminosity function is a growing function with magnitude, we expect more objects with $m_{meas} < m_{threshold}$, but $m_{true} > m_{threshold}$ than objects with $m_{meas} > m_{threshold}$ and $m_{true} < m_{threshold}$. So, in our measured sample we have more galaxies than in the sample selected with the true magnitude with the same magnitude threshold, making our selection wider, and therefore, reducing the total power.

### 6.5 2-point correlation function

Using the samples presented in previous sections, we measure the 2-point correlation function using the package `TreeCorr` (Jarvis et al. 2004) with the Landy & Szalay estimator (Landy & Szalay 1993).

$$w(\theta) = \frac{DD - 2DR + RR}{RR} \tag{3}$$

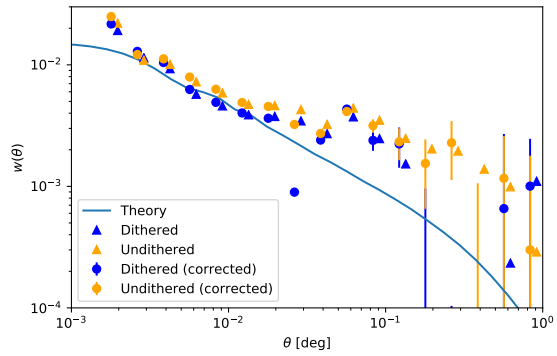


**Figure 17.** Results for the two-point correlation function in the input (green), dithered (blue) and undithered (orange) datasets. We can see that after the correction (triangles) for the systematic effects the agreement is better than without the correction (circles) between the input and output data especially for the undithered case (for the dithered case the corrections are negligible).

where $DD, DR$, and $RR$ are the number of pairs of objects taking from the data $D$ or the random catalog $R$ that covers the footprint. We use 30 angular log-spaced bins between $\theta = 0.0001°$ and $\theta = 10°$. The number of bins has been chosen so that the resulting covariance matrix is nearly diagonal. The covariance matrices are calculated using the delete-one jackknife technique (Shao 1986; Norberg et al. 2009). We divide the footprint in $N_{JK} = 100$ regions. These regions are defined using the K-means algorithm from the package `kmeans_radec`[8]. The covariance matrix is then computed as

$$\mathrm{Cov}_{JK}(\theta_i, \theta_j) = \frac{N_{JK} - 1}{N_{JK}} \sum_{k=1}^{N_{JK}} \Delta w_k(\theta_i) \Delta w_k(\theta_j) \tag{4}$$

$$\Delta w_k(\theta_i) = w_k(\theta_i) - \bar{w}(\theta_i) \tag{5}$$

Where $w_k(\theta_i)$ is the value of the correlation function when deleting the $k$-th region at the scale $\theta_i$, and $\bar{w}(\theta_i)$ is the average correlation function at that same scale. We compute the correlation function on the dithered and undithered catalogs using their respective footprints, and we compare with the predicted correlation function given the $N(z)$ obtained by the spatial matching presented in previous sections, and using `CCL`[9] (LSST Dark Energy Science Collaboration et al. 2018). The results can be seen at Figure 17.

#### 6.5.1 *Angular power spectrum*

We measure the angular power spectrum using the `NaMaster`[10] package (Alonso et al. 2018). This package computes the cross-power-spectra of masked fields with an arbitrary number of contaminants using a pseudo-$C_\ell$ approach (Hivon et al. 2002; Elsner et al. 2017). As we do for real space, we select the number of $\ell$ bins so that the covariance matrix is almost diagonal. In this case we calculate the power spectrum in the range $0 < \ell < 6144$ and

[8] https://github.com/esheldon/kmeans_radec
[9] https://github.com/LSSTDESC/CCL
[10] https://github.com/damonge/NaMaster

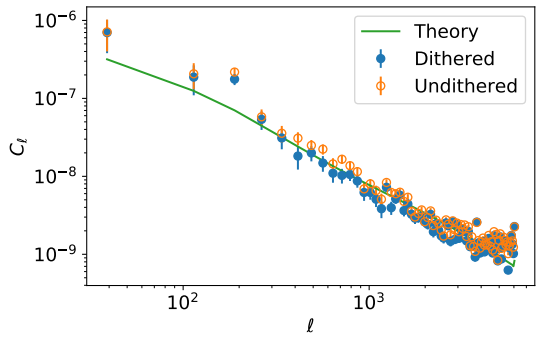


**Figure 18.** Measured power spectra undithered (open orange circles) and dithered (solid blue circles) datasets with `NaMaster` corrected by systematics. The error bars are computed using the Gaussian approximation. All datasets seem to be compatible with the theoretical prediction (green line) within 1-$\sigma$.

$\Delta\ell = 75$. We calculate the covariance matrices with three different approaches: One consists in the same delete-one jackknife technique performed in real space; in the other approach we compute the covariances using the mode-counting formula (Dodelson 2003; Cabré et al. 2007)

$$\mathrm{Cov}_{\ell\ell'} = \frac{2}{f_{sky}\Delta\ell}\left(\frac{C_\ell^2}{2\ell+1} + \frac{1}{\bar{n}^2}\right)\delta_{\ell\ell'} \tag{6}$$

where $\bar{n}$ is the number density (objects per steradian). The last approach we compute the Gaussian covariance with `NaMaster`. The three methods give consistent results. We compute the theoretical prediction for the power-spectra with `CCL`:

$$C_\ell^{TH} = \frac{2}{\pi}\int dz \left(\frac{dn(z)}{dz}\right)^2 b^2(z) \int dk k^2 P(k,z) j_\ell^2(kr(z)) \tag{7}$$

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 $\frac{dn}{dz}$ built with the input catalog using the same magnitude cuts. The results can be seen in Figure 18. We can appreciate that the results for both datasets seem to follow the theoretical prediction within errors.

### *6.5.2 Systematic effects*

In this section we analyze the different systematic effects affecting the DC1 data. We will consider the following observational quantities as sources for systematic uncertainties:

- Extinction: The CatSim catalog provides the value for the magnitudes already corrected for extinction using the SFD map (Schlegel et al. 1998). We use this map to cross-correlate with our galaxy catalogs.
- Stellar contamination: In this case we just build a HEALPix map using the input CatSim stellar catalog to cross-correlate with our galaxy catalogs.
- Sky-background: We use the observed background level in each exposure and assign that value to the HEALPixel with $N_{side} = 2048$ that corresponds to the pointing position. After this we calculate the median value in each HEALPixel and use them to compute the cross-correlations with the galaxy catalogs. The caveat of this approach is that we are not propagating the geometry of the focal plane.
- Sky-noise: We use the observed noise background level in each exposure and proceed as in the previous case to build a HEALPix map with which we will cross-correlate.
- Seeing: We proceed as before and use the observed seeing in each exposure and build a HEALPix map.

These maps are shown in Figure A1 and Figure A2.

In the case of the real space measurements, we proceed the same way as ref. (Crocce et al. 2016) to compute the impact of the different potential sources for systematic uncertainty. We compute the auto and cross-correlations of these maps and our data samples to obtain the "true" correlation function:

$$w^{gg}(\theta)_{true} = w^{gg}(\theta)_{obs} - \vec{w}_{g,sys} \cdot W^{-1}_{sys,sys} \cdot \vec{w}_{g,sys} \tag{8}$$

Where $w^{gg}_{true}$ is the corrected galaxy-galaxy correlation function, $w^{gg}_{obs}$ is the measured galaxy-galaxy autocorrelation, $\vec{w}_{g,sys}$ is the vector containing the cross-correlation between the galaxies and the different maps and $W^{-1}_{sys,sys}$ is the inverse of the matrix containing the cross-correlations between different systematics. For the stellar contamination we also follow the procedure presented in ref. (Crocce et al. 2016). Given a stellar fraction $f_{star}$ the "true" galaxy correlation function is given by

$$w_{gal} = (1+f_{star})^2\left(w_{obs} - f_{star}^2 w_{sg} - \frac{f_{star}^4}{(1+f_{star})^2}\right) \tag{9}$$

where $w_{sg}$ is the cross-correlation between our stellar map and the observed galaxies.

Comparing with the input catalog we select those objects classified as galaxies whose centroids lie within 2 pixels of a star from the input catalog. From those, we select the objects that have a magnitude difference smaller than 30 mmags. Doing this we estimate a stellar contamination of $f_{star} = 7.2\%$ in our sample.

We compare the correction term for the different systematics to the measured signal obtaining the results in Figure 19, where we can appreciate that the dithering strategy is working to reduce the impact of the systematics, keeping them under control and at percent level for $\theta < 0.03 degrees$ in this "small" area, improving the signal-to-noise in addition to other benefits explored in ref. (Awan et al. 2016). This is also seen Figure 17 where we clearly see that the uncorrected and corrected correlation function for the dithered field are essentially the same.

In the case of harmonic space we use the mode deprojection from `NaMaster` (Alonso et al. 2018) to correct for the different templates. In Figure 20 we show the relative size of the correction due to the presence of the considered sources of systematic uncertainty. Again, we see that the dithering strategy diminishes the effect of the potential sources of systematic uncertainty in the power-spectrum measurements.

## 7 CONCLUSIONS

End-to-end simulations are very useful tools to test the overall performance of any current and future cosmological experiments like the LSST (Ivezić et al. 2008). They allow us

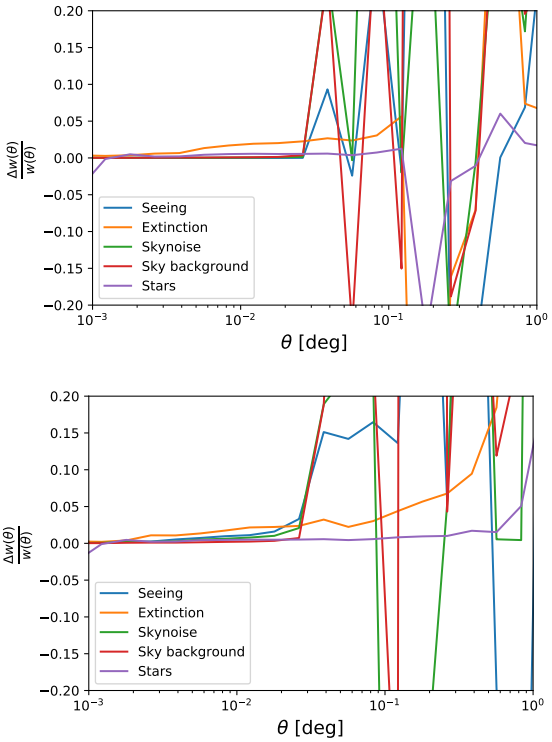


**Figure 19.** Correction due to the different potential sources of systematic uncertainty relative to the value of the measured correlation function of the dithered (top) and undithered (bottom) datasets. We show that the correction is at percent level for $\theta < 0.03$ degrees and grows beyond the 20% level for $\theta \approx 0.1$ degrees.

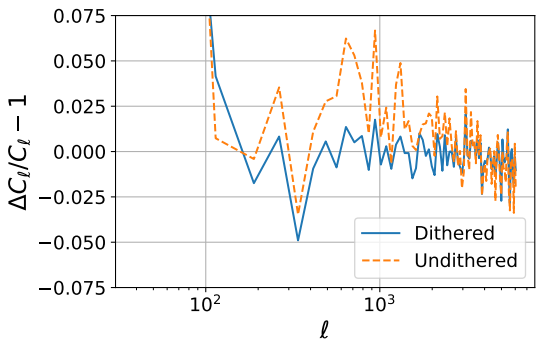


**Figure 20.** Correction in harmonic space due to the different potential sources of systematic uncertainty relative to the value of the measured correlation function of the dithered (solid blue) and undithered (dashed orange) datasets.

to test and improve different parts involved in the data processing and analysis. On the other hand, they also allow us to model and improve our control of systematic uncertainties.

In this paper we have presented a simulated (end-to-end) imaging dataset to resemble LSST data, corresponding to the first data challenge in 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 dithering (*undithered*). Then we processed these images with the LSST DM stack. We started by performing several quality assurance tests on the outputs from the DM stack, including photometry and astrometry checks. We checked that both the *dithered* and *undithered* are high-quality datasets with good photometry and astrometry (with the caveat of uncorrected proper motion).

After this, we studied different ways to relate the output catalogs to the inputs. In particular, we analyzed spatial matching, and spatial matching including magnitude matching. We saw that the angular power-spectrum is recovered at a very high-level of precision for both cases, however, these matching techniques are likely not good enough for studies about blending or very small scale information. One of the main advantages of the end-to-end simulations is that photons can be traced back to the original sources, however, this is not always possible due different factors such as, data volume limitations so, an important topic for these kind of end-to-end simulations will be to find efficient strategies to relate inputs and outputs.

Finally, we estimated the depth and selected a high-completeness sample to perform clustering analysis, in both real and harmonic space. The results of this analysis, indicate that the simulated foregrounds have a low impact at the scales considered for our study (lower than 5% for $\theta < 0.03°$ and for the $\ell$ range studied in harmonic space) especially, in the *dithered* dataset. This indicates the success of the dithering strategy considered in this study.

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

## Acknowledgments

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.

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

## REFERENCES

Alonso, D., Sánchez, J., Slosar, A., et al. 2018, in prep.
Awan, H., Gawiser, E., Kurczynski, P., et al. 2016, ApJ, 829, 50
Bosch, J., Armstrong, R., Bickerton, S., et al. 2017, ArXiv e-prints, arXiv:1705.06766
Bruderer, C., Chang, C., Refregier, A., et al. 2016, ApJ, 817, 25
Brun, R., Hagelberg, R., Hansroul, M., & Lassalle, J. C. 1978, Simulation program for particle physics experiments, GEANT: user guide and reference manual (Geneva: CERN)
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Cabré, A., Fosalba, P., Gaztañaga, E., & Manera, M. 2007, MNRAS, 381, 1347

Crocce, M., Carretero, J., Bauer, A. H., et al. 2016, MNRAS, 455, 4301
De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, ArXiv e-prints, arXiv:1611.00036
Dodelson, S. 2003, Modern cosmology (San Diego, CA: Academic Press)
Elsner, F., Leistedt, B., & Peiris, H. V. 2017, MNRAS, 465, 1847
Gonzalez, O. A., Rejkuba, M., Zoccali, M., Valenti, E., & Minniti, D. 2011, A&A, 534, A3
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Hirata, C., & Seljak, U. 2003, MNRAS, 343, 459
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Ivezić, Z., Tyson, J. A., Abel, B., et al. 2008, ArXiv e-prints, arXiv:0805.2366
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864
Jurić, M., Kantor, J., Lim, K., et al. 2015, ArXiv e-prints, arXiv:1512.07914
Krisciunas, K., & Schaefer, B. E. 1991, PASP, 103, 1033
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
LSST Dark Energy Science Collaboration. 2012, ArXiv e-prints, arXiv:1211.0310
LSST Dark Energy Science Collaboration, Chisari, N. E., Alonso, D., et al. 2018, in prep.
LSST Science Collaboration. 2009, ArXiv e-prints, arXiv:0912.0201
Mandelbaum, R., Hirata, C. M., Seljak, U., et al. 2005, MNRAS, 361, 1287
Mckinney, W. 2011
Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Peterson, J. R., Jernigan, J. G., Kahn, S. M., et al. 2015, ApJS, 218, 14
Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Shao, J. 1986, Annals of Statistics, 14, 1322
Sjöstrand, T., Mrenna, S., & Skands, P. 2006, Journal of High Energy Physics, 5, 026
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629 EP
Suchyta, E., Huff, E. M., Aleksić, J., et al. 2016, MNRAS, 457, 786


## APPENDIX A: MAPPING OBSERVATIONAL EFFECTS

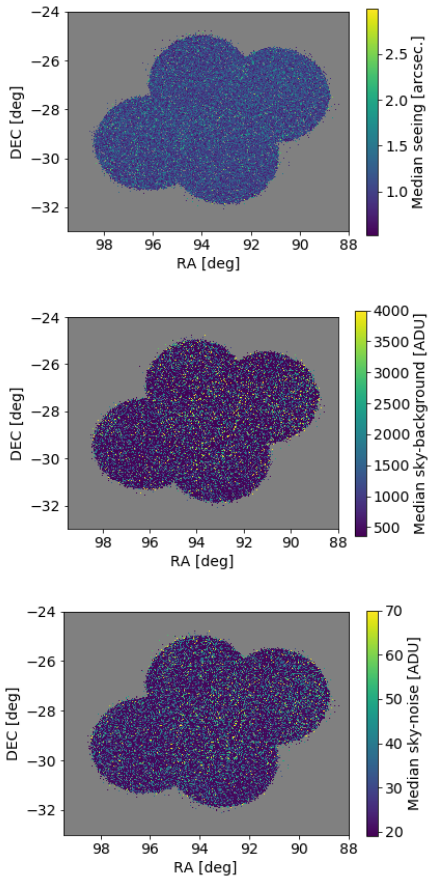


**Figure A1.** HEALPix maps showing the different observational effects that might be potential cause of systematic uncertanties.

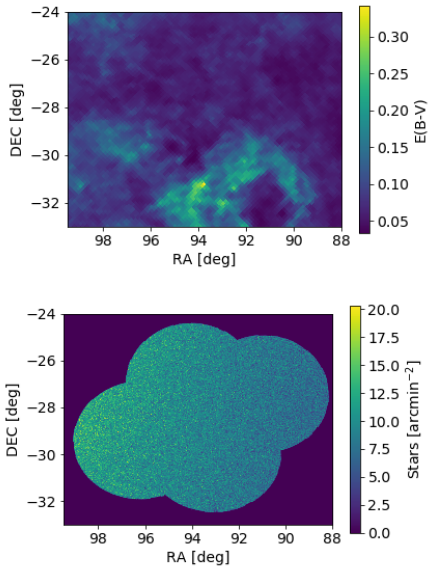


**Figure A2.** HEALPix maps showing the different observational effects that might be potential cause of systematic uncertanties.