

LARGE SYNOPTIC SURVEY TELESCOPE

## Large Synoptic Survey Telescope (LSST) Data Management

# Options for Photometric Redshifts for the LSST Data Release Object Catalog

M. L. Graham, J. Bosch, L. P. Guy, (ADD MORE), and the DM System  
Science Team.

DMTN-049

Latest Revision: 2019-08-30

DRAFT

### Abstract

This document presents and discusses options for photometric redshifts for the LSST Data Release Objects catalog.

## Change Record

Version	Date	Description	Owner name
1	2017-04-01	Initial release of preliminary investigation.	Melissa Graham
2	2018-10-16	Edited to align with recent DPDD updates, some of which were based on the recommendations of Version 1 of this document.	Melissa Graham
3	2019-XX-XX	Updated as per ticket/DM-6367.	Melissa Graham

## Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 Use-Cases for Object Photo-<math>z</math></b>	<b>2</b>
2.1 Internal DMS Use-Cases . . . . .	2
2.2 Scientific Use-Cases . . . . .	2
2.3 Considerations for LSST Year 1 . . . . .	4
<b>3 Proposed "Minimum Viable Product" for Object Photo-<math>z</math></b>	<b>4</b>
<b>4 Proposed Pathway to Generating Object Photo-<math>z</math></b>	<b>5</b>
4.1 DM-SST Questions (and Proposed Answers) . . . . .	6
4.2 The Option to Just Ingest a Community Photo- $z$ Catalog After the DR . . . . .	7
4.2.1 Photo- $z$ As An In-Kind Contribution for Data Rights . . . . .	8
<b>5 Needed LSST Data for Training/Calibration</b>	<b>8</b>
<b>6 Storing Object Photo-<math>z</math></b>	<b>9</b>
<b>7 Object Catalog Values That Enable User Generated Photo-<math>z</math></b>	<b>10</b>
<b>8 Proposed Diagnostics for Object Catalog Photo-<math>z</math></b>	<b>11</b>
<b>9 Proposed Selection Method for Photo-<math>z</math> Algorithms</b>	<b>12</b>
9.1 Lessons from Work Comparing Photo- $z$ Estimators . . . . .	12
9.2 An Example Comparative Analysis of Photo- $z$ Codes for LSST Photometry . . . . .	13
9.2.1 Simulated Catalogs . . . . .	13
9.2.2 Considered Photo- $z$ Estimators . . . . .	14
9.2.3 Analysis and Results of the Comparison . . . . .	15

# Options for Photometric Redshifts for the LSST Data Release Object Catalog

## 1 Introduction

A *photometric redshift* is an estimate of an object's cosmological redshift (distance) which is based on its photometry (the apparent magnitudes and colors obtained in multiple filters) instead of on its spectral features (such as emission and absorption lines). Cosmological redshifts are a key component of the LSST's science goals, but it is currently impossible to obtain spectra for the billions of galaxies that LSST will observe. Thus, photometric redshifts for the LSST Object catalog will be a key component of a wide variety of LSST science.

Typically, photometric redshift algorithms either fit template spectra to the observed photometry or match photometry to a training set of galaxies with spectroscopic redshifts. The latter is often done with machine learning codes, and hybrid photo- $z$  estimators also exist. Some photo- $z$  estimators are more appropriate for some science goals than others, due to the quality or type of results they produce, and for this reason several research groups in the science community are already planning to generate their own photo- $z$ .

There is a requirement on the Data Management System (DMS) which states that "*The DMS shall compute a photometric redshift for all detected Objects*" [LSE-61]. This document proposes to define the *minimum viable product* for the Object photo- $z$  as meeting the *basic* science needs for communities which will not or cannot generate custom photo- $z$  (Section 2 and 3). Research and development of photometric redshift algorithms for LSST is beyond the scope of DM (not part of DM's specialized knowledge base), as it is itself an active area of current and future LSST research; instead, one or more existing photo- $z$  algorithm(s) would be installed by DM into the DR pipeline. The purpose of this document is to propose a path for the decision of which estimator(s) to use, and for the implementation of photo- $z$  into the DR pipelines (Section 4). Additional components and supporting data for the generation of Object photo- $z$  are also presented and discussed.

DMS-REQ-0046

## 2 Use-Cases for Object Photo- $z$

Some of the following information on use-cases was collected from participants of the LSST Project and Community Workshop’s session on Photometric Redshifts on Aug 14 2019<sup>1</sup>.

### 2.1 Internal DMS Use-Cases

Since DM’s galaxy photometry outputs are being developed with the goal of feeding photometric redshift algorithms, and it is expected that computing photometric redshifts will be a part of the science validation process. Unlike stars, color-color and color-magnitude diagrams for galaxies do not have sufficient structure to reveal issues with the photometry. While other photometric validation techniques will also be useful (such as evaluating the width of galaxy cluster red-sequences) they may only apply to *some* galaxies, whereas *all* galaxies have a redshift.

The internal use-case of scientifically validating the galaxy photometry outputs is likely to require a simple photo- $z$  estimator which fits SED templates, since the goal is to evaluate whether the photometric outputs match the colors of real galaxies. Whether this simpler photo- $z$  could also serve the scientific use-cases is undetermined, because the photometric validation process is not yet defined or written.

As a side note, although the DMS will assign fiducial spectral energy distributions (SEDs) to Objects in order to apply sub-band wavelength-dependent photometric calibration and PSF modeling, computing photo- $z$  is not planned to be a part of this process. Furthermore, the SED templates used will likely be simpler (e.g., step-function or slope) than would be needed for deriving photo- $z$ .

### 2.2 Scientific Use-Cases

**Dark Energy** – Extragalactic astrophysics such as weak lensing, baryon acoustic oscillations, and Type Ia supernova cosmology are all main science drivers for the LSST, and all require catalogs of galaxies with photometric redshifts. The photo- $z$  algorithms for precision cosmology will be custom-tailored to these particular science goals. For example, weak lensing and large

<sup>1</sup>Thanks to Sam Schmidt, Chris Morrison, Sugata Kaviraj, Gautham Narayan, Lauren Corlies, Travis Rector, Tina Peters, Alex Malz, and other participants.

scale structure require ensemble measurements of  $N(z)$ , whereas point-estimate photo- $z$  for individual objects are required for Type Ia supernova host galaxies and the identification of strong lensing candidates and galaxy cluster members. The Dark Energy Science Collaboration is developing specialized photo- $z$  pipelines for these science goals (which could serve to generate photo- $z$  for the object catalog, as discussed in Section ??).

**Galaxies** – The Galaxies Science Collaboration reported that they would use LSST-provided object photo- $z$ , and that their science goals require that photo- $z$  be accurate enough (< 10%) to derive intrinsic galaxy properties like mass and star formation rate. They also indicated that posteriors delivered as  $P(z, M)$  and/or with rest-frame apparent magnitudes would be useful to their science goals; this indicates that the results of a template-fitting photo- $z$  estimator might be more relevant to Galaxies studies than machine-learning estimates. The object photo- $z$  might also be used to assist with star-galaxy separation, to enable population studies, to estimate environmental (clustering) parameters, and/or to choose instrument configurations for spectroscopic follow-up (i.e., the expected location of emission lines).

**Time Domain** – The Transients and Variable Stars Science Collaboration reported that they would use LSST-provided object photo- $z$  to identify and/or characterize extragalactic transient host galaxies. Alert packets provide object IDs for the three nearest stars and three nearest galaxies in the most recent data release. Alert stream brokers intend to query the objects catalog in real time to obtain host photo- $z$  because photometric classification for transient light curves is *significantly* aided by redshift estimates. The object catalog's photo- $z$  will also be used to identify and prioritize the potential host galaxies of gravitational wave events for imaging searches of the optical counterpart.

**Stars, Milky Way, and Local Volume** – LSST-provided object photo- $z$  could be used to reject compact extragalactic objects from stellar samples for population studies and/or spectroscopic follow-up campaigns.

**Active Galactic Nuclei** – It is currently unclear how useful the object photo- $z$  will be for the AGN community because there is no special deblending planned for the DMS to produce galaxy photometry which is free of AGN emission. The AGN contribution to the DR CoAdd image stacks, and thus the object catalog photometry, will be an average flux over the LSST survey images. Photometric redshift codes will either have to be able to recognize and deal with AGN contamination, or the photo- $z$  estimates for AGN host galaxies will be impacted. Potential AGN contamination could be identified by identifying DIAObjects in the nuclear region,

but quantifying and removing that AGN flux from the galaxy photometry and recalculating photo- $z$  remain a user-generated data product.

**Education and Public Outreach** – The question “*how far away is it?*” is common to many EPO initiatives and the Object catalog photo- $z$  will be used when preparing information for the public. EPO might also use photo- $z$  for, e.g., generating 3D graphics that visualize large volumes, or educational programs on the Hubble constant. For EPO purposes, high precision is not as important as outlier reduction for photo- $z$ .

## 2.3 Considerations for LSST Year 1

To maximize early science capabilities, algorithms that will return the most accurate photo- $z$  as early in the survey as possible could be prioritized. In the first year of LSST, it might be simpler to use a template-fitting photo- $z$  estimator and avoid potential issues related to computation resources and/or the need to train a machine learning model. Additionally, the large spectroscopic training sets needed for ML photo- $z$  estimators are more likely to exist by 2030 than at 2020. However, if a machine learning estimator is applied for LSST DR 1 and 2, it should be a community-accepted algorithm with demonstrated success in other surveys, preferably surveys that overlap the LSST volume, as this will facilitate the characterization and validation of the LSST photo- $z$ .

## 3 Proposed “Minimum Viable Product” for Object Photo- $z$

Based on the scientific use-cases for Object photo- $z$  presented in Section 2, the following is proposed for the type, basic attributes, and minimum performance of the Object photo- $z$  in order for it to meet the *the basic science needs* for communities which will not or cannot generate custom photo- $z$ .

**The following must be refined by iterating with the scientific community.**

### Type (Template-fit or Machine Learned)

Descriptions of the internal and Galaxies science use-cases indicate that a template-fitting photo- $z$  estimator, as opposed to an estimator based on machine learning, might be preferred in those cases. Thus it is proposed that Object catalog photo- $z$  should be based on a template-fitting algorithm *and then also* a machine-learning algorithm, if multiple results can

be stored (see Section 6). The best fit template should be easily accessible by users.

## Attributes

To serve all use-cases, the LSST Object catalog photo- $z$  results should be easily understandable, well characterized, and documented. The Object catalog photo- $z$  should have reliable uncertainties and/or flags to help the novice user avoid mis-applying the results, or overestimating their significance. The Object catalog photo- $z$  should always be populated, starting from the time of data release. The main science driver of this basic attribute is alert brokers, which require host-galaxy photo- $z$  to optimally classify and prioritize transients for follow-up, and obtain this information via each alert's associations with nearby Objects from *the most recent DR*. (Otherwise, an older DR would have to be associated to new DIASources for alerts until the new DR's Object photo- $z$  are ready.)

## Performance

Generally, it seems that the non-DESC use-cases for object photometric redshifts can tolerate results with a point-estimate accuracy of  $\sim 10\%$ . Thus it is proposed that a photo- $z$  estimator which can return results with a standard deviation in  $z_{\text{true}} - z_{\text{phot}}$  of  $\sigma_z < 0.05(1 + z_{\text{phot}})$ , and a catastrophic outlier fraction of  $f_{\text{outlier}} < 10\%$ , over a redshift range of  $0.0 < z_{\text{phot}} < 2.0$ , seems likely to serve the broader community well. It is proposed that these quality goals would apply to  $i < 25$  mag galaxies.

## 4 Proposed Pathway to Generating Object Photo- $z$

### #0. Define the minimum viable product.

Project and community work together to refine the type, basic attributes, and minimum performance of Object catalog photo- $z$  to meet *the basic* science needs for communities which will not or cannot generate custom photo- $z$ . A proposed definition for the minimum viable product is put forth in Section 3. Preliminarily, it seems very likely that off-the-shelf photo- $z$  estimators will be able to fulfill these basic necessities.

### #1. Define the selection criteria and asses algorithm(s).

Project and community work together to refine the algorithm selection criteria and evaluate the currently available photo- $z$  algorithms. A preliminary proposal for selection criteria and potential algorithms is put forth in Section 9.

## #2. Select the algorithm(s).

DM chooses one or more algorithms to implement, based on the jointly established evaluation criteria *and internal considerations such as computational resources*.

## #3. Implement, calibrate, and/or train the algorithm.

DM implements the algorithm(s) into the DR pipeline, along with any needed supporting data sets such as templates or spec- $z$  catalogs, and diagnostic code (see #5). The community is welcomed to participate in calibration/training with "DR previews" released shortly before (weeks/months) a new DR.

## #4. Run the algorithm, populate the Object catalog.

DM runs the algorithm as part of DR processing. The Object catalog is populated with photo- $z$  results prior to DR.

## #5. Provide diagnostics.

Diagnostics regarding the photo- $z$  results are made available (e.g., as auto-generated documentation); potential diagnostic tests are collected in Section 8. The community is welcomed and encouraged to participate in this workflow item.

The above step should *not* be construed as "science validation of the photo- $z$ ", because that is beyond scope: these photo- $z$  are not intended to meet all science needs. DM would provide these diagnostics to inform users' application of the Object catalog photo- $z$ , but validation of these photo- $z$  for individual science goals is left to the users.

## #6. Upgrade the Object table with a superior community product if possible.

The DM system as delivered to the Operations Project provides a minimal scientific capability with respect to Object catalog photo- $z$ , but if that is rendered obsolete by community efforts, then the superior product should be ingested and federated.

### 4.1 DM-SST Questions (and Proposed Answers)

#### How much work will it take DM to generate Object photo- $z$ ?

The proposed workflow requires DM to engage with the community now to define the minimum viable product and the criteria for selecting a photo- $z$  algorithm. This might require, e.g., circulating this document, ingesting feedback, and participating in Science Collaboration telecons or face-to-face meetings. The bulk of DM's work will be in implementing the selected

algorithm into the DR pipeline, which will also include the installation of template or training sets, and setting up photo- $z$  diagnostic codes to inform community use of the Object photo- $z$ . This process will probably “take as much work as”, e.g., implementing treatments for crowded fields, calibrating star/galaxy separation, or training the real/bogus classifier for DIASources.

The proposal put forth to (1) define a minimum viable product for Object photo- $z$  and (2) not commit to a full science validation reduces the potential of photo- $z$  generation to overrun DM budgets *and* makes it more likely that off-the-shelf photo- $z$  estimators will suffice.

### **How should this work be distributed between project and community?**

Given that photometric redshifts are a required project deliverable, the project should not require that the community participate or rely on any community-generated products. However, since they are the ultimate end-user of the photo- $z$  the community should be welcomed and encouraged to participate (as described in the proposed itemized pathway). Should the community generate a superior photo- $z$  catalog and agree to release it to the data rights community, it should be federated and replace the project-derived photo- $z$  in the Object catalog.

### **Will Object photo- $z$ be a condition of data release?**

Yes. They are a defined deliverable of DM, and scientifically necessary. For example, having Object photo- $z$  at the time of data release is likely to be especially important for alert brokers, because DIASources will be associated with Objects from the most recent data release. This is how brokers will obtain photo- $z$  estimates for transient host galaxies, which is an essential component of classification and prioritization for follow-up.

### **What is the amount of DR processing resources that can be dedicated to one or more photometric redshift algorithm(s) and/or compression (Section 6).**

TBD.

### **When and how might DM generate and use Object photo- $z$ internally for, e.g., photometric validation? Would these photo- $z$ also meet the scientific use-cases (Section 2.1)?**

TBD.

## **4.2 The Option to Just Ingest a Community Photo- $z$ Catalog After the DR**

If this option were to become the primary plan for generating Object photo- $z$ , then the steps for the project and the community would be as follows.

1. Well before data release, the science community submits proposals to generate photometric redshifts for the Object catalog.
2. Before the data release, the project selects one or more of the proposals. The project works with the community team to ensure the requisite LSST calibration data will be obtained of, e.g., spec- $z$  deep fields.
3. Shortly before data release, the project provides the community team(s) with a "DR preview" so that they can train and calibrate their photo- $z$  algorithm.
4. After data release, the community team generates Object catalog photometric redshifts, performs quality assessment and validation, and prepares the associated documentation that would be needed for user applications.
5. After data release, the project ingests and federates the photometric redshifts into the Object catalog.

Although this option would be significantly less work for the DM/Ops teams, there are many potential **drawbacks**:

- no initiative or reward for the community team which has generated the photo- $z$  (except maybe citations to their photo- $z$  catalog)
- risk of having no Object photo- $z$  if no community team delivers
- risk of having Object photo- $z$  which are tailored to the specific science case of the community team and do not serve the broader science use-cases
- a DR without photo- $z$  is a problem for brokers (unless alerts are associated to an older DR's Object catalog)
- does not satisfy the requirement that the "*DMS shall compute*" photo- $z$

#### 4.2.1 Photo- $z$ As An In-Kind Contribution for Data Rights

Any plan which requires an MoU or similar arrangement between the LSST Operations Project and members of the community would have to be agreed on with, e.g., Bob Blum. Such a plan cannot be negotiated or decided on by DM, or by any aspect of the construction project alone, although DM (or, e.g., the PST) may suggest or endorse a plan.

## 5 Needed LSST Data for Training/Calibration

**The following should be refined by iterating with the scientific community.**

Regardless of the type or process for generating the Object photo- $z$ , some training and calibration data from LSST will be needed. For example, deep multi-band LSST imaging and photometry for spectroscopic fields like COSMOS, and WFD-level imaging that overlaps, e.g., DESI/4MOST, obtained during commissioning in order to produce photo- $z$  in year 1 (if possible).

Some photo- $z$  methods have requirements other than spec- $z$  fields: e.g., Sánchez & Bernstein [17] use clustering information to obtain photo- $z$  and this requires wider, shallower field coverage and not a single deep pointing like a spec- $z$  field would have. This wider area would also serve to reduce cosmic variance in the training set (~100 square degrees would serve to average out the variance).

For the community to participate in the training or calibration of a photo- $z$  algorithm prior to a data release, it might be necessary to release a small (~ 10%) but representative DR "preview" data set. The community – especially the DESC PZ working group and the Galaxies SC – should be included in the process to generate a list of fields which would be best to include in such a DR "preview".

## 6 Storing Object Photo- $z$

**The following could be refined by iterating with the scientific community.**

The LSST Data Products Definitions Document (DPDD) [LSE-163] defines the format of the Object catalog's table columns which could store the results of photometric redshift estimates, regardless of how they're generated. The following is from Table 5 of the DPDD:

- `photoZ` (`float[2x95]`) = photometric redshift likelihood samples – pairs of redshift and likelihood ( $z, \log L$ ) – computed using a to-be-determined published and widely accepted algorithm at the time of LSST Commissioning
- `photoZ_pest` (`float[10]`) = point estimates for the photometric redshift provided in `photoZ`

The exact point estimate quantities stored in the `photoZ_pest` are to-be-determined, "*but likely candidates are the mode, mean, standard deviation, skewness, kurtosis, and 1%, 5%, 25%, 50%, 75%, and 99% points from cumulative distribution*" [LSE-163]. Both the posteriors and point estimates from several different photo- $z$  estimators could be compressed and stored in this

allotted space, as discussed below.

Given the variety of use-cases and the fact that different photo- $z$  estimators produce different results (Schmidt et al., in prep.), the option to compute, compress, and store estimates from multiple algorithms in the  $2 \times 95$  float might be scientifically desirable. Efficient  $P(z)$  compression algorithms are in development, such as Carrasco Kind & Brunner [6] and Malz et al. [15]. Carrasco Kind & Brunner [6] present an algorithm for sparse representation, for which “*an entire PDF can be stored by using a 4-byte integer per basis function*” and “*only ten to twenty points per galaxy are sufficient to reconstruct both the individual PDFs and the ensemble redshift distribution,  $N(z)$ , to an accuracy of 99.9% when compared to the one built using the original PDFs computed with a resolution of  $\delta z = 0.01$ , reducing the required storage of two hundred original values by a factor of ten to twenty.*” Malz et al. [15] presents a Python package for compressing one-dimensional posterior distribution functions (PDFs), demonstrates its performance on several types of photo- $z$  PDFs, and provides a set of recommendations for best practices which should be consulted when DM is making decisions on the DR photo- $z$  data products.

However, compression (and decompression by users) will require extra computational resources, which should be estimated and considered. Decompression must be fast and easy for users to implement.

## 7 Object Catalog Values That Enable User Generated Photo- $z$

**The following could be refined by iterating with the scientific community.**

It is important to ensure that the photometric parameters computed and provided in the Object table are sufficient for users to generate photometric redshifts from custom codes. Aside from fluxes and/or apparent magnitudes and errors for each LSST filter, this might include the color properties in the Object table [LSE-163]:

- stdColor (float[5]) = ‘standard color’, color of the object measured in ‘standard seeing’, suitable for photo- $z$
- stdColorErr (float[5]) = uncertainty on stdColor

As described in LSE-163: “*Colors of the object in ‘standard seeing’ (for example, the third quartile expected survey seeing in the  $i$  band,  $\sim 0.9$  arcsec) will be measured. These colors are guaranteed to be seeing-insensitive, suitable for estimation of photometric redshifts.*”

Another quantity that is useful for photo- $z$  is the effective transmission function ( $\phi$ ; Eq. 5 in LPM-17), which will be provided for all Sources either in the catalog or as a link.

## 8 Proposed Diagnostics for Object Catalog Photo- $z$

**The following must be refined by iterating with the scientific community.**

These diagnostics would be based on the attributes/performance defined in Section 3 and the evaluation criteria defined in Section 9. These could be automatically generated for the Object photo- $z$  (like OpSim metrics).

Based on the LSST Object catalog alone, some diagnostics could be:

- comparing the sum of the  $P(z)$  to the ensemble  $N(z)$
- smoothness of  $N(z)$  and  $P(z)$  (or at their expected shape)
- how  $z_{\text{phot}}$  uncertainty correlates with photometric error
- check for discrepancy between star/galaxy and photo- $z$

Based on a cross-match between the LSST Object catalog to other spec- $z$  and/or phot- $z$  catalogs (e.g., Zhou et al. 20), some diagnostics could be:

- comparing phot/spec distributions  $N(z)$  for a set of color or magnitude cuts
- compare the shapes of  $P(z)$  for individual objects
- spec. vs. phot. redshift plots and associated statistics (standard deviation, bias, fraction of outliers)

DESC is currently investigating methods of validating accuracy of probability distributions from a photo- $z$  algorithm (see e.g., Schmidt, Malz et al. 2019 in prep.) and those methods could be used to generate diagnostics.

The scientific community may want to perform their own advanced diagnostics to inform their use of the LSST photo- $z$ , such as:

- assessing the performance of point estimates in broker photometric classification algorithms
- evaluating the absolute magnitude distribution of Type Ia supernovae once the distance derived from the photo- $z$  point estimates are applied
- evaluating the distributions of derived physical parameters for galaxies using the  $P(z)$

- checking whether high SFR galaxies (and maybe other sensitive populations) have reasonable  $P(z)$
- galaxy cluster membership identification

## 9 Proposed Selection Method for Photo- $z$ Algorithms

The full selection criteria should include aspects such as performance, attributes, ease of use and implementation, the existence of publicly available training or template sets, demonstrated success for other wide-field optical surveys (and overlap of those surveys with the future LSST area), and community adoption.

This is an extremely preliminary demonstration of the types of tests a given photo- $z$  estimator might undergo during the selection process. In Section 9.1 we cover some existing work that compares photo- $z$  estimators. In Section 9.2 we apply a couple of off-the-shelf estimators to simulated LSST photometry to demonstrate a few statistical techniques for comparing photometric redshifts. Since this is mainly for *demonstrative* purposes, the estimators themselves have not been optimized to return the most accurate photo- $z$ , and some minor mistakes have been left uncorrected.

**Pretty much all of the following can be replaced with a citation to Schmidt et al. (2019) when that paper is ready.**

### 9.1 Lessons from Work Comparing Photo- $z$ Estimators

**DESC photo- $z$  WG** – This science community is fully engaged in the development of photo- $z$  routines and their optimization for LSST; their work is not reproduced here.

**Relevant photo- $z$  testing papers** – [10] tested 18 different photo- $z$  codes on the same sets of simulated and real data and found no significantly outstanding method. [7] test 11 different photo- $z$  codes on the CANDLES data set ( $U$ -band through infrared) and also find that no method stands out as the “best,” and that there is a strong dependence of photo- $z$  accuracy on the SNR of the photometry (relevant for our tests at 1 year). They also found that most of the photo- $z$  codes underestimate their redshift errors, which is important to note because we do want accurate errors.

**Lessons from DES** – [18] use the science verification data (200 square degrees of  $grizY$  photometry to a depth of  $i_{AB} = 24$  magnitudes) of the Dark Energy Survey (DES) to evaluate several photometric redshift estimators. They found that the Trees for Photo- $z$  code (TPZ; Carrasco Kind & Brunner 4) provided the most accurate results with the highest redshift resolution, and that template-fitting methods also performed well – especially with priors – but that in general there was no clear “winner.”

**Lessons from SDSS** – [1] describes the photo- $z$  adopted for the SDSS DR12. They first use an empirical technique with a large training set to estimate the redshift and its error, and then fit SED templates with that redshift in order to obtain additional galaxy information such as  $K$ -correction and spectral type. (Note: *They call it a hybrid technique, but the photo- $z$  sounds like it comes solely from the local linear regression, basically an interpolation in the color-redshift relation.*)

**Lessons from Lucy’s work** – Summer student Lucy Halperin (UW 2016, with Melissa Graham) took what we call the “Brown” catalog<sup>2</sup> made by Sam Schmidt, with simulated 10-year LSST-like magnitude uncertainties, and ran it through 2 machine learning (ANNz and TPZ) and 2 template-fitting (LePhare and BPZ) photo- $z$  codes. All four returned sets of photo- $z$  with similar standard deviations and biases, but the template-fitting codes were more prone to failures and outliers. Lucy’s work found that for template-fitting photo- $z$  codes, the choice of template SED set does make a significant difference in the results, particularly regarding photo- $z$  outliers – however, this may have been particular to the use of the “Brown”-based galaxy catalog.

## 9.2 An Example Comparative Analysis of Photo- $z$ Codes for LSST Photometry

We have used simulated galaxy catalogs (Section ??) and three different photo- $z$  estimators (Section ??) to generate photometric redshifts. We then apply a series of diagnostics to evaluate and compare their performance (Section ??).

### 9.2.1 Simulated Catalogs

We use a randomly chosen 30000 galaxy test subset of the LC\_DEEP\_GONZALEZ2014A catalog, which is based on the Millennium simulation [19] and the galaxy formation models of

<sup>2</sup>We call it the Brown catalog because it uses the SEDs from [3]

[9] and constructed using the lightcone techniques described by [16]. We impose a limit on the true catalog redshift of  $z < 3.5$ , and a limit on the apparent  $i$ -band magnitude of  $i < 25.5$ , and furthermore require galaxies to be detected in the three filters  $gri$ . The latter requirement means that the test galaxies' apparent magnitude is brighter than a limit defined by a signal-to-noise ratio  $< 5$  in all three filters  $gri$ . This limit depends on the number of years of survey elapsed, and since we want to use the same set of test galaxies to analyze the algorithms' results early in the survey, we require this  $gri$  non-detection with the expected limits after only 1 year of LSST. These restrictions mean that we end up with a catalog that has with fewer faint galaxies than will be in the LSST 10-year catalogs, and so the 10-year results we consider here are optimistic (but that's fine for our purposes). These restrictions are imposed prior to the random selection of 30000 test galaxies from the larger catalog. We then simulate 4 versions of the test galaxy catalog with errors appropriate for 1, 2, 5, and 10 years of LSST. We calculate galaxy magnitude uncertainties that are appropriate for the elapsed survey time, and observed photometry is simulated by adding a random scatter proportional to the uncertainties.

In addition to the test set, we need a training set of galaxies for the machine-learning algorithm to serve as a spectroscopic redshift catalog. Spectroscopic data sets containing tens of thousands of galaxies down to  $i > 25$  and out to  $z > 3$  are certainly possible, e.g., the VIMOS Ultra Deep Survey (VUDS; Le Fèvre et al. 14). Assuming that the LSST will cover a spectroscopic field like the VUDS to the full 10-year depth during commissioning or with a first-year deep drilling field, we use as our training set a sample of 30000 catalog galaxies with photometric uncertainties equivalent to a 10-year LSST. This training set has the same redshift and magnitude distribution and limits as the galaxy catalogs, which may not be the case for a real spectroscopic set.

### 9.2.2 Considered Photo- $z$ Estimators

Here we've considered one template-fitting and one machine-learning photo- $z$  algorithm. Hybrid photo- $z$  estimators attempt to mitigate the flaws of either process (e.g., the SDSS DR12 photo- $z$  estimator by Beck et al. [1], or the Gaussian Processes estimator described by [?]).

**Bayesian Photometric Redshifts** (BPZ; Benítez 2) is a template-fitting algorithm with a magnitude prior<sup>3</sup>. We use all default parameters, including the  $i$ -band for the magnitude prior,

---

<sup>3</sup><http://www.stsci.edu/~dcoe/BPZ/>

except that we supply the CFHTLS set of SED templates. This set is 66 SEDs that were used for the CFHTLS photo- $z$  paper and are from [11], and they were interpolated from the CWW and Kinney models.

**Trees for Photometric Redshifts** (TPZ; Carrasco Kind & Brunner 4, 5) is a machine learning algorithm that uses prediction trees and a training set of galaxies with known redshifts. We use all the default parameters from the example, except we increase the number of trees from 4 to 10 (this was set low in the provided example to decrease run time). Since the number of realizations is 2, this is a total of 20 trees. As shown in [5], the bias and scatter of the resulting photo- $z$  improve the most as the number of trees is increased to 20, and continues to improve more mildly to 100, and then are not much improved beyond 100 trees (i.e., their Figure 9). We also set the maximum redshift to 3.5 and the number of redshift bins to 350. We include both magnitudes and colors and their uncertainties as attributes to be used in the prediction trees, as Lucy's work found that this led to better results. From the TPZ output files, we take as  $z_{\text{phot}}$  the mode of the redshift distribution instead of the mean because this is the peak of the distribution (most likely redshift). *Note: I may have misunderstood how TPZ uses the photometric uncertainties. I thought it treated errors differently from other Attributes, but perhaps it uses them just the same. That makes sense, but means it is inappropriate to include photometric errors as ttc Attributes if the train and test sets have different photometric precision.* In case a training set with LSST photometric uncertainties at the level of a 10-year survey is not available from commissioning or a dedicated deep drilling survey by the end of year 1, we also simulate the photo- $z$  results with a training set that has the same level of photometric uncertainty as the test set.

**Color-Matched Nearest-Neighbors** (CMNN; Graham et al. 2018) is a photo- $z$  estimator that uses the Mahalanobis distance in color-space to match a galaxy to a training set. We simulate photo- $z$  at 1, 2, 5, and 10 years using a test set of 20000 and a training set of 60000. Training set has the same photometric depth as the test set.

### 9.2.3 Analysis and Results of the Comparison

In this section we demonstrate several analysis techniques for comparing the output of different photo- $z$  estimators, discussing each in turn below.

#### The $z_{\text{true}} - z_{\text{phot}}$ Diagram

Figure 1 shows the photo- $z$  results in  $z_{\text{true}} - z_{\text{phot}}$  diagrams, which are a typical way to visually

assesses the output. Galaxies are plotted with a semi-transparent black dot so that density and clustering of points is clear, and galaxies that end up designated as “outliers” are over-plotted with a more opaque red dot. Problems such as outlier structure from e.g., color- $z$  degeneracies, and quantization of  $z_{\text{phot}}$  is obvious in these kinds of diagrams, as well as a decent overall impression of the scatter and bias. Even though our runs with these estimators have not been optimized, we can make an “example” assessment of these figures to compare the two estimators. **BPZ:** Overall the results are quite poor even with the LSST 10-year predicted photometry, especially the amount of quantization at  $z_{\text{phot}} > 1.5$ . We find that the results are not improved if we remove the magnitude prior, or use a different SED template set such as those from [3]. We do know from Lucy’s work that the choice of SED template set has a significant impact on the results (this is actually widely known) – in Lucy’s work, the best results were achieved when we used the Brown SEDs with a galaxy catalog for which the photometry was simulated using those same SEDs. However, it’s less straightforward to identify the “best” template SEDs to use with the Euclid galaxy catalog (or real data for that matter). [9] describes the wide variety of stellar population spectral synthesis models they used, but it would take quite some work to get them all together into a single catalog to provide to BPZ. **TPZ:** The results are quite poor 1 and 2 years, with a lot of quantization in the photo- $z$  and many outliers at low and high redshift, but are significantly improved at 10 years. It is very interesting that there is actually a large improvement if the training set does not have better photometric errors than the test set (i.e., compare the 1 and 2 year results in the third row to the second row), but this may just be related to how we’ve included the errors as Attributes. Either way, TPZ is sensitive to the provided training set, so an extended investigation into what would truly be a realistic 1 year spectroscopic training set for LSST should be done (e.g., different redshift distributions, different magnitude limits). Although TPZ appears to give better accuracy, we also need to ensure that it gives realistic precision for its photo- $z$  results.

## Statistical Measures

The important statistical measures that are typically used to assess photo- $z$  results are based on the photo- $z$  error,  $\Delta z_{(1+z)} = (z_{\text{spec}} - z_{\text{phot}})/(1 + z_{\text{phot}})$ . We measure the robust standard deviation in  $\Delta z_{(1+z)}$ ,  $\sigma_{\Delta z_{(1+z)}}$  (i.e., “robust” because it is the standard deviation of galaxies within the IQR); the robust bias, which is the mean deviation  $\bar{\Delta z}_{(1+z)}$ ; and the fraction of outliers,  $f_{\text{out}}$ , which is the fraction of galaxies with  $|\Delta z_{(1+z)}| > 0.06$  and  $> 3\sigma_{\text{IQR}}$  (i.e., must be greater than whichever constraint is larger). In the community sometimes the median deviation in  $\Delta z_{(1+z)}$  over all galaxies is used instead of the mean deviation of galaxies within the IQR, but we find the two are comparable. In Figure 2 we demonstrate a convenient way to statistically compare the results from multiple photo- $z$  estimators. In this case we are comparing the values of

these statistical measures when the photo- $z$  estimators are run on galaxy catalogs simulated to represent the 1, 2, 5, and 10 year DRP from LSST (colored lines), for both BPZ (left) and TPZ (right). Different estimators for the same year could also be plotted in a single graph. From these statistical measures it is obvious, for example, that the photo- $z$  from TPZ outperform those from BPZ at all years. In Figure 3 we show examples of how to compare the statistical measures for the full catalog (i.e.,  $0.3 \leq z_{\text{phot}} \leq 3.0$ ) for different photo- $z$  estimators by plotting, e.g., the fractions of failures versus the outliers, or the bias versus the standard deviation.

### **Photo- $z$ Uncertainties, $\delta z_{\text{phot}}$**

In Figure 4 we demonstrate a way to assess the photo- $z$  uncertainties,  $\delta z_{\text{phot}}$ , that come out of the estimators: we plot  $z_{\text{phot}} - \delta z_{\text{phot}}$  in the main axis, and above and to the side plot the distributions in  $\delta z_{\text{phot}}$ ,  $z_{\text{phot}}$ , and for comparison,  $z_{\text{true}}$ . With BPZ, we can see a strict floor in the photo- $z$  uncertainty that increases with redshift (i.e., the uncertainties are bogus, though this could be a fault of mine in running the code and not of the code itself). For both BPZ and TPZ we can see that in some cases the clumps causing a quantization in photo- $z$  also have high photo- $z$  uncertainty, suggesting that a simple cut on  $\delta z_{\text{phot}}$  could return a sample for which the photo- $z$  distribution matches the true distribution. However, there are other clumps in photo- $z$  that have a relatively low uncertainty. Overall, from these plots we could conclude that the TPZ algorithm returns a redshift distribution that is more similar to the true distribution. Another option here is to plot the photo- $z$  error ( $\Delta z_{(1+z)}$ ).

### **The Posterior Probability Density Function, $P(z)$**

In Figure 5 we plot examples of the posterior probability density functions output by the BPZ and TPZ algorithms for two test galaxies. One galaxy was chosen as a random representative of galaxies for which an inaccurate and imprecise photo- $z$  was returned from both BPZ and TPZ for all years (top panel of Figure 5). The other was chosen as a random representative of galaxies which experienced a large and consistent improvement in both the accuracy and precision of its photo- $z$  from year 1 to 10, for both BPZ and TPZ (bottom panel of Figure 5). These kind of plots demonstrate, for example, the quantization in the TPZ photo- $z$  in the PDFs (this may be related to a mistake in the TPZ input).

### **Q-Q Plots**

In Figure 6 we show an example of a quantile-quantile plot using the true *v.s.* the photometric redshift. Each point represents the  $z_{\text{true}}$  and  $z_{\text{phot}}$  for a given quantile, and since the two distributions we are comparing have the same total number of objects (we've neglected any galaxies that have failed to return a photo- $z$ ), we're simply using  $1/N$  as the quantiles. If the

Q-Q plot is linear with a slope of 1, we would know the distributions of  $z_{\text{phot}}$  would match that of  $z_{\text{true}}$ . In Figure 6 we can see for both BPZ and TPZ that this is not the case, but that BPZ is worse.

Draft

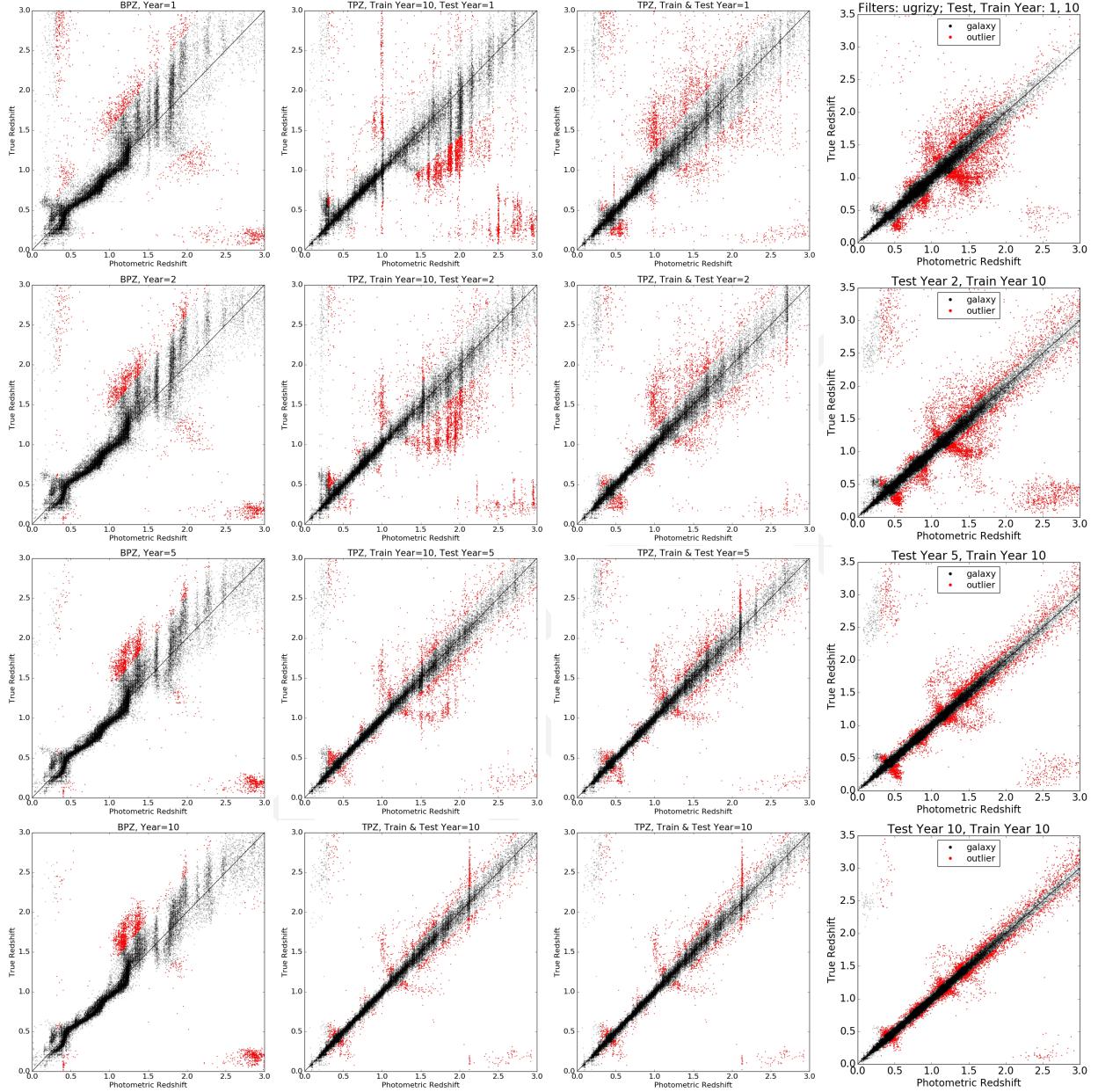


FIGURE 1: Examples of  $z_{\text{true}} - z_{\text{phot}}$  plots for a variety of algorithms (by column), for 1, 2, 5, or 10 years of survey time elapsed (top to bottom). Galaxies that are statistical outliers are shown in red. **Left:** results for the BPZ algorithm. **Center-left:** results for the TPZ algorithm with a 10-year training set. **Center-right:** results for the TPZ algorithm with a co-evolving training set. **Right:** results for nearest-neighbors color-matching algorithm, with 50000 test galaxies and  $10^6$  training set galaxies with co-evolving photometric errors.

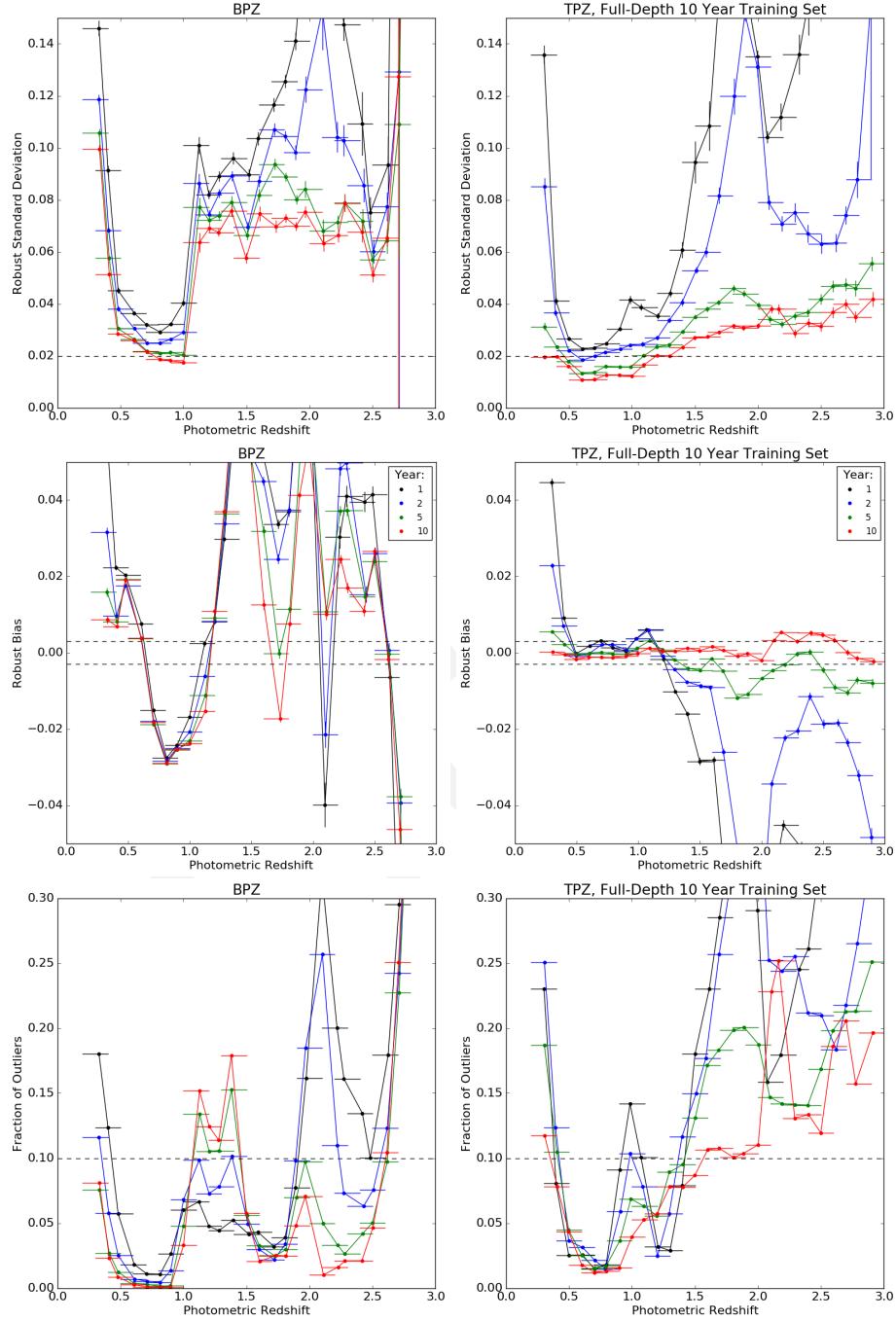


FIGURE 2: Examples of a statistical measures of the photo- $z$  results from BPZ (left) and TPZ with an evolving training set (right) for simulated catalogs at 1 to 10 years (line colors as in plot legends). From top to bottom we show the robust standard deviation from the IQR, the robust bias, and the fraction of outliers as a function of photo- $z$ , with matched  $x$ - and  $y$ -axes to facilitate comparison between BPZ and TPZ.

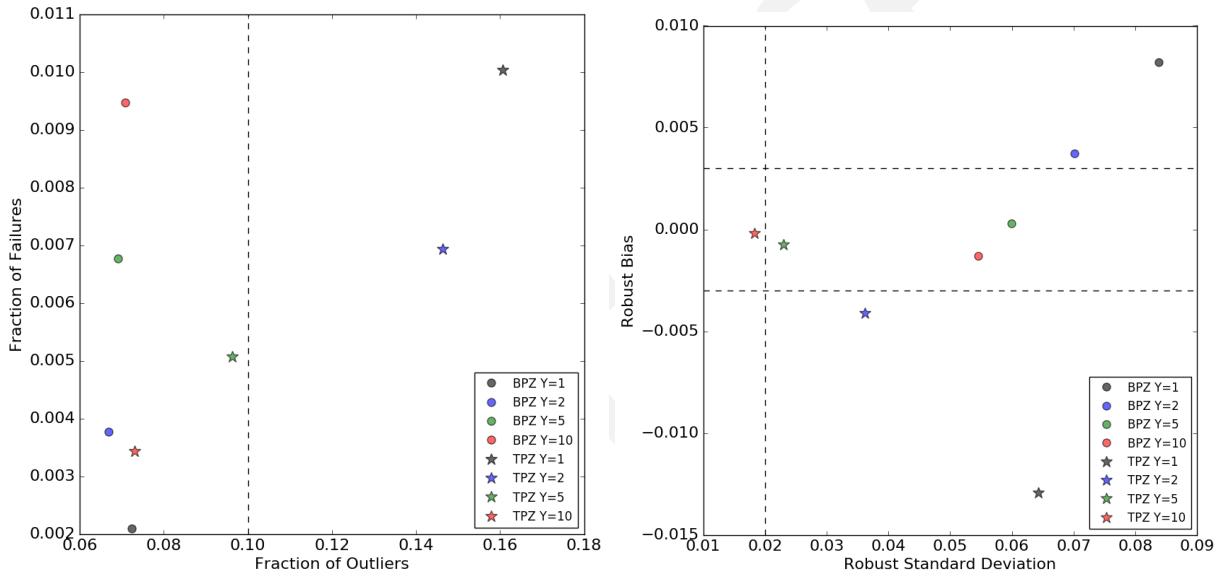


FIGURE 3: Examples of how to compare statistical measures over  $0.3 \leq z_{\text{phot}} \leq 3.0$  from different photo- $z$  estimators by plotting one against the other: fraction of failures and outliers (left), and the robust bias and standard deviation (right). In this case we're comparing the statistical measures for TPZ and BPZ from photometry simulated for the LSST at years 1, 2, 5, and 10 (legend).

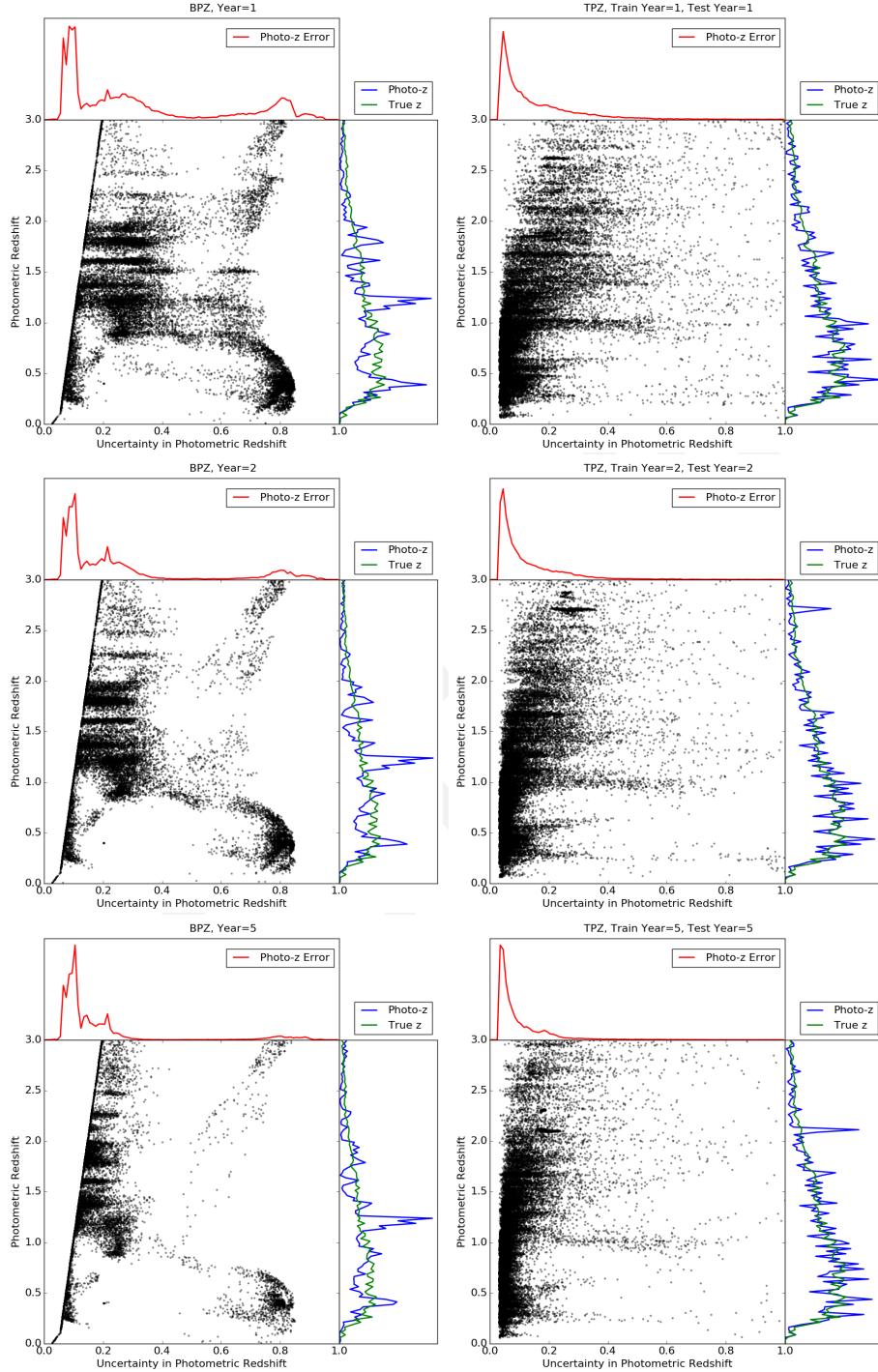


FIGURE 4: Examples of plot to compare the photo- $z$  uncertainty ( $\delta z_{\text{phot}}$ ) between algorithms with  $z_{\text{phot}} - \delta z_{\text{phot}}$  plots from the BPZ (left) and TPZ (right) estimators for simulated catalogs with photometric uncertainties at 1, 2, and 5 years of LSST (top to bottom). Red lines show the distribution of photo- $z$  errors; blue and green lines compare the distributions of true and photometric redshifts.

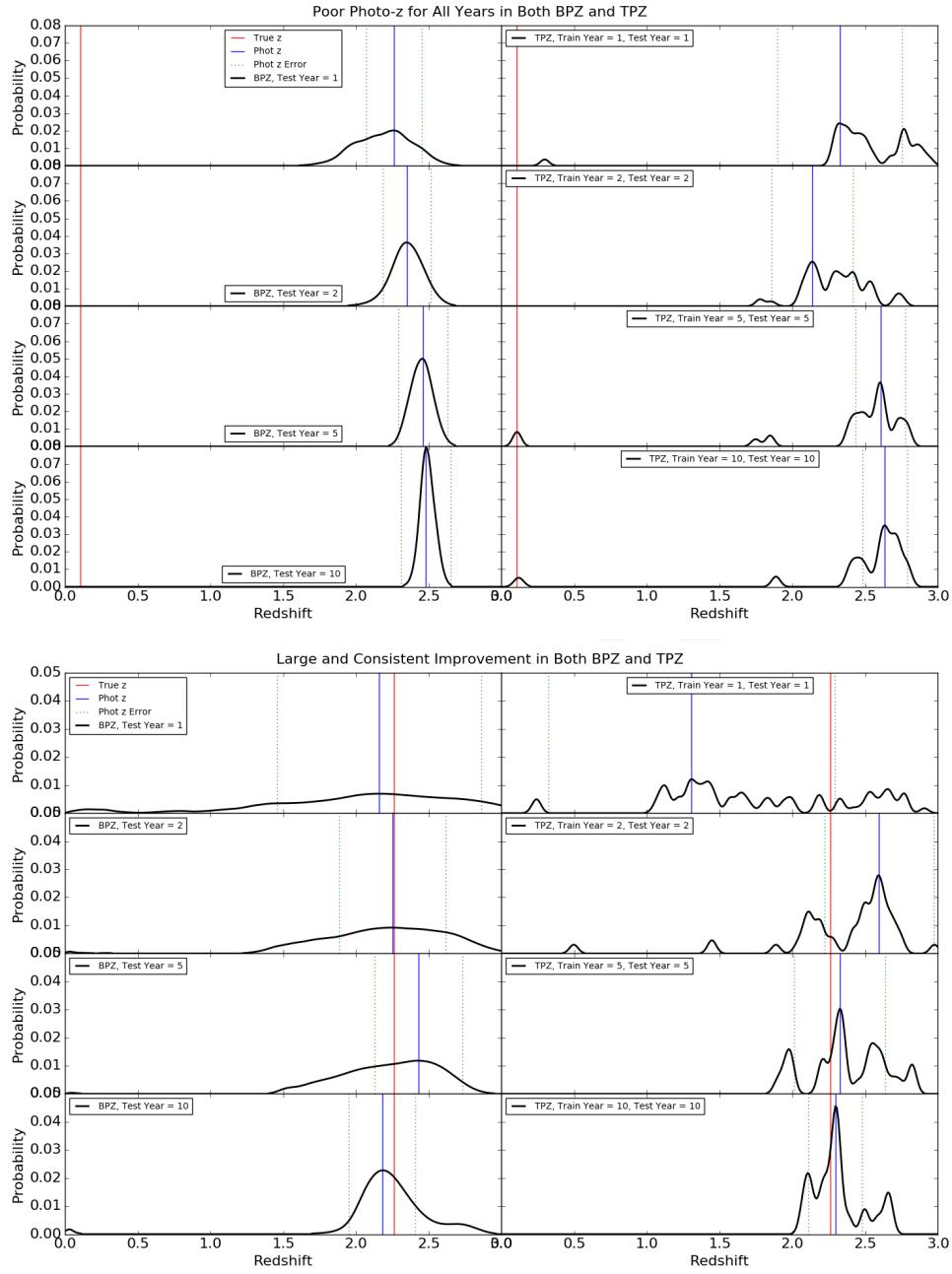


FIGURE 5: Examples of the posterior probability density functions for two test galaxies in all of our simulations: BPZ (left) and TPZ (right) for photometric uncertainties like 1, 2, 5, and 10 years of LSST (rows from top to bottom). In the top panel we choose a galaxy that return inaccurate and imprecise photo-z from all 8 trials, and in the bottom panel we choose a galaxy that experienced a large and consistent improvement in photo-z accuracy and precision from 1 to 10 years with both estimators.

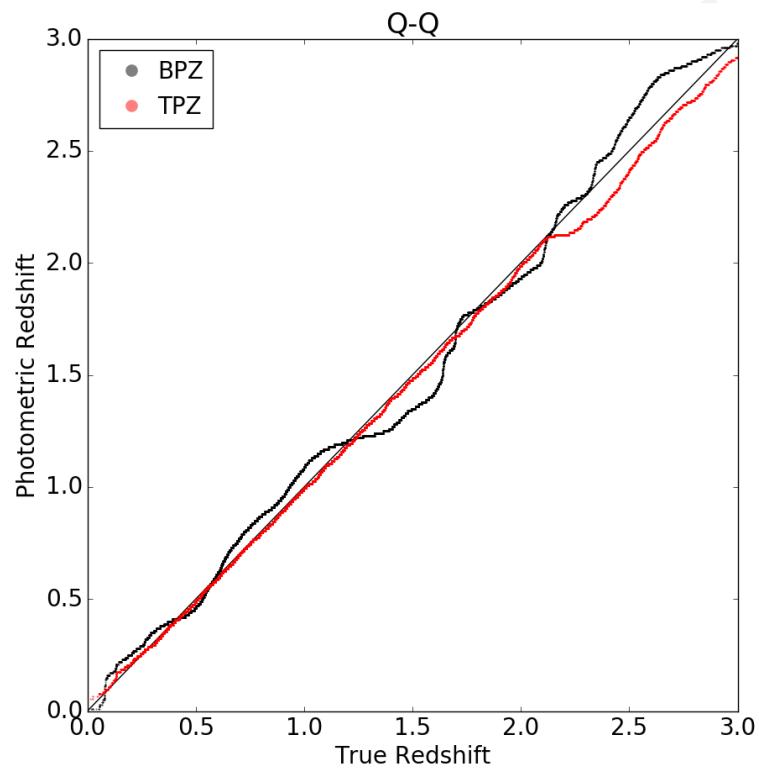


FIGURE 6: Example of a Q-Q plot, using  $z_{\text{true}}$  and  $z_{\text{phot}}$ , from the BPZ and TPZ estimators.

## References

- [1] Beck, R., Dobos, L., Budavári, T., Szalay, A.S., Csabai, I., 2016, MNRAS, 460, 1371 (arXiv:1603.09708), doi:10.1093/mnras/stw1009, ADS Link
- [2] Benítez, N., 2000, ApJ, 536, 571 (arXiv:astro-ph/9811189), doi:10.1086/308947, ADS Link
- [3] Brown, M.J.I., Moustakas, J., Smith, J.D.T., et al., 2014, ApJS, 212, 18 (arXiv:1312.3029), doi:10.1088/0067-0049/212/2/18, ADS Link
- [4] Carrasco Kind, M., Brunner, R., 2013, TPZ: Trees for Photo-Z, Astrophysics Source Code Library (ascl:1304.011), ADS Link
- [5] Carrasco Kind, M., Brunner, R.J., 2013, MNRAS, 432, 1483 (arXiv:1303.7269), doi:10.1093/mnras/stt574, ADS Link
- [6] Carrasco Kind, M., Brunner, R.J., 2014, MNRAS, 441, 3550 (arXiv:1404.6442), doi:10.1093/mnras/stu827, ADS Link
- [7] Dahlen, T., Mobasher, B., Faber, S.M., et al., 2013, ApJ, 775, 93 (arXiv:1308.5353), doi:10.1088/0004-637X/775/2/93, ADS Link
- [8] **[LSE-61]**, Dubois-Felsmann, G., Jenness, T., 2018, *LSST Data Management Subsystem Requirements*, LSE-61, URL <https://ls.st/LSE-61>
- [9] Gonzalez-Perez, V., Lacey, C.G., Baugh, C.M., et al., 2014, MNRAS, 439, 264 (arXiv:1309.7057), doi:10.1093/mnras/stt2410, ADS Link
- [10] Hildebrandt, H., Arnouts, S., Capak, P., et al., 2010, A&A, 523, A31 (arXiv:1008.0658), doi:10.1051/0004-6361/201014885, ADS Link
- [11] Ilbert, O., Arnouts, S., McCracken, H.J., et al., 2006, A&A, 457, 841 (arXiv:astro-ph/0603217), doi:10.1051/0004-6361:20065138, ADS Link
- [12] **[LPM-17]**, Ivezić, Ž., The LSST Science Collaboration, 2018, *LSST Science Requirements Document*, LPM-17, URL <https://ls.st/LPM-17>
- [13] **[LSE-163]**, Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL <https://ls.st/LSE-163>
- [14] Le Fèvre, O., Tasca, L.A.M., Cassata, P., et al., 2015, A&A, 576, A79 (arXiv:1403.3938), doi:10.1051/0004-6361/201423829, ADS Link

- [15] Malz, A.I., Marshall, P.J., DeRose, J., et al., 2018, AJ, 156, 35 (arXiv:1806.00014), doi:10.3847/1538-3881/aac6b5, ADS Link
- [16] Merson, A.I., Baugh, C.M., Helly, J.C., et al., 2013, MNRAS, 429, 556 (arXiv:1206.4049), doi:10.1093/mnras/sts355, ADS Link
- [17] Sánchez, C., Bernstein, G.M., 2019, MNRAS, 483, 2801 (arXiv:1807.11873), doi:10.1093/mnras/sty3222, ADS Link
- [18] Sánchez, C., Carrasco Kind, M., Lin, H., et al., 2014, MNRAS, 445, 1482 (arXiv:1406.4407), doi:10.1093/mnras/stu1836, ADS Link
- [19] Springel, V., White, S.D.M., Jenkins, A., et al., 2005, Nature, 435, 629 (arXiv:astro-ph/0504097), doi:10.1038/nature03597, ADS Link
- [20] Zhou, R., Cooper, M.C., Newman, J.A., et al., 2019, MNRAS, 488, 4565 (arXiv:1903.08174), doi:10.1093/mnras/stz1866, ADS Link