LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST)

LSST DRP (Level 2)
Catalog Photometric Redshifts

M. L. Graham

DMTN-???

Latest Revision: 2017-05-03

revision: TBD
status: draft

Abstract

The purpose of this document is to begin to assemble the diversity of motivations driving the inclusion of photometric redshifts in the LSST Level 2 Object Catalog, and prepare to make a decision on what kind of photo-z products will be used. The roadmap for this process is described in Section 1. We consider the photo-z use-cases in order to validate that the type of photo-z incorporated into the Level 2 DRP catalog, and the format in which it is stored, meets the needs of both DM and the community. We also compile potential evaluation methods for photo-z algorithms, and demonstrate these options by applying them to the photo-z results of two off-the-shelf photo-z estimators. The long-term plan is for this document to develop over time and eventually describe the decision-making process and the details of the selected algorithm(s) and products. **Preliminary recommendations can be found in Section 1.**

Change Record

Version	Date	Description	Owner name
1	2017-04-??	Initial release of preliminary investigation.	Melissa Graham

Contents

1   Introduction

1.1   Roadmap for LSST Level 2 DRP Photo-z

This roadmap is directly from comments by Mario Jurić on the Jira thread[1].

The process for the definition of the L2 photo-z data product is the same as for any other data product that isn't fully defined yet. The canonical steps are:

(1) DM Project Science (acting on behalf of overall Project Science) consults with the community (represented by the relevant collaboration – in this case, the DESC) for options and proposals as to what form of the data product is needed, and the selection of the appropriate algorithm. While it is likely that we will adopt what the DESC recommends, it's likely there will be some iterating to converge to a scientifically acceptable yet implementable solution (e.g. the recommended algorithm may run afoul of our computational, storage, or budget limitations; also DESC may not be the only group concerned about the photo-z product and its properties).

(2) Following that (iterative) consultation, DM will recommend (and the Project will select) the Photo-Z algorithm and the data product format, for use in Level 2.

(3) DM will implement the selected algorithm. We will, of course, use whatever we can transfer over from DESC (or other) work, but our requirements may be higher than DESC's (e.g., we need to run it reliably in LSST's production environment).

(4) DM will implement anything needed to integrate, QC, QA the photo-z data product in data releases (and everything leading up to those; e.g., commissioning). We will, of course, make use of any code and/or help from the DESC, but it is the Project deliverable.

1.2   Preliminary Recommendations

(1) DM should write down their plans regarding whether and how the `Object` database's photo-z will be used in the calibration and/or QA processes, and/or if there will be a separate, internal-use photo-z for this purpose, perhaps in a different database. (Section 2.1)

(2) DM should update the DPDD to include the fact that each `Object.Source` will have a link to its own unique filter transmission function (ϕ, Eq. 5 in the SRD); this is an important quantity for photo-z and the SC need to know of it's existence and format. (Section 3)

(3) Related to above, DM should establish a timeline for developing the algorithm (or releas-

---

[1]If you have permissions, access at `https://jira.lsstcorp.org/browse/DM-6367`

ing this algorithm, if it is already in place) that will be used to standardize the effective filter curves when generating the CoAdds, and verifying how well it meets specifications (e.g., systematics ~ 0.005 magnitudes). (Section 3)

(4) DM should expand the number of database elements allotted to photo-z from just the `float[2x100]` that will contain the $P(z)$ by adding one `float[4]` for a "best" photometric redshift and it's uncertainty (mean, median, 68%, 95%), and one e.g., `float[1]`, to encode the best-fit SED template (or synthesized mix of templates) when applicable. However, using 5 of the allocated 200 floats would probably be a suitable compromise. (Section 3)

(5) DM should establish a timeline for this process and propose a deadline for the decision regarding the process for generating the photo-z and how they will be stored in the Level 2 database.

(6) DM should continue to interact with the Science Collaborations to validate that the specifications related to the Level 2 DRP catalog's photo-z meet their needs. One option may be to officially request the report from the Pitt photo-z meeting which is in preparation and would provide useful input.

2   Use-Cases as Motivation for the Level 2 Photo-z Estimator(s)

In this section we describe the internal and external motivations for why photo-z are being included in the Level 2 database and how they influence our choice of photo-z algorithm. Typically, photometric redshift algorithms work by either fitting template spectra to the observed photometry or by matching photometry to a training set of galaxies with spectroscopic redshifts. The method of "matching photometry" is often done with machine learning codes. Hybrid photo-z estimators also exist, attempting 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 Leistedt & Hogg 12). To know which kind of photo-z estimator(s) LSST will need, we start by collecting the various motivations for including a photo-z estimate in the Level 2 catalogs.

2.1   Internal Requirements

The relevant requirement is DMS-REQ-0046, "The DMS shall compute a photometric redshift for all detected Objects". From the Jira conversation on this issue we've managed to cobble together some potential internal uses for a Level 2 photo-z: **A)** "DM would like to be able to run some sort of photo-z code for QA purposes ... good enough to validate LSST's galaxy

photometry". **B)** DM "should definitely also run some sort of photometric cluster finder for QA purposes, which provides much better photo-zs for red cluster galaxies and hence gives us much better photometric quality tests for a small subset of the population ... idea is that the width of the red sequence in clusters is really the only simple metric we have for galaxy photometry colors that doesn't rely on simulations". However, photo-z might not actually be involved in this metric, as the code uses "the colors of the galaxies directly to estimate the red sequence, rather than relying on photo-zs (which have a lot of assumptions built in already)". **C)** "DM is already on the hook for some sort of photo-z in the form of SED fitting (and I would assume templates/redshifts) for chromatic corrections".

While vague, all of these seem to suggest at least implicitly that a template-fitting code would be used to generate the photo-z. **A)** Using observed SED of real galaxies seems like the best choice for validating photometry, e.g., if the question is whether the photometric outputs match the colors of real galaxies. **B)** If photo-z for red sequence galaxies are being prioritized, perhaps the best results would be generated by fitting elliptical galaxy spectra only. Also, we'd need to test whether template-fitting or machine-learning is best for that particular scenario. **C)** In this case, it sounds like SED fitting is the only option that would satisfy the use case of chromatic corrections.

Reconciling the photo-z option that best suits internal needs with external might be quite difficult and furthermore, unnecessary. Jeff Newman from DESC-PZ commented that the internal-use photo-z *"will almost certainly be different/simpler photo-zs than the database ones".*

Recommendation: DM should write down their plans regarding whether and how the `Object` database's photo-z will be used in the calibration and/or QA processes, and/or if there will be a separate, internal-use photo-z for this purpose, perhaps in a different databse.

2.2 External Requirements

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. However, the photo-z algorithms for precision cosmology will be custom-tailored to the needs of each domain and we should consider those applications as Level 3 photometric redshifts. Instead, the main consumers of the Level 2 photo-z product will probably be users wanting to do things like:

- evaluate the cluster membership status of galaxies (e.g., for spectroscopic follow-up campaigns)
- reject compact extragalactic objects from stellar samples, or vice versa (e.g., for spectroscopic follow-up campaigns)
- identify and/or characterize extragalactic transient host galaxies (sometimes the transient redshift is known and used to identify the host, and sometimes the most probable host is used to constrain the transient redshift)
- begin their cosmological studies as soon as possible, before the Level 3 photometric redshift routines are perfected
- additionally some advanced/specialized science may start with the Level 2 photo-z and then apply calibrations, adjustments, or cuts on top of that in order to use them (e.g., maybe weak lensing studies; from a conversation with H. Hildebrandt)

The chosen Level 2 photo-z algorithm(s) should be easily understandable and characterizable, and the output photometric redshifts and their uncertainties should be strictly defined. If a tradeoff needs to be made, it will probably be more important for the uncertainties to be reliable than for the photo-z to be more accurate (to a point, of course). To maximize early science capabilities of the scientific community we should consider algorithms that will return the most accurate photo-z as early in the survey as possible. For this reason, in Section 4 we begin to set up a framework for a performance analysis of photo-z from galaxy catalogs simulated to resemble LSST data at 1, 2, 5 and 10 years.

In the future we will be be in touch with the extragalactic LSST Science Collaborations for input (and we are already in touch with the DESC photo-z WG). We will do this by contacting the chairs of following SC: Galaxies, Dark Energy, Active Galactic Nuclei, Transients, and Strong Lensing (unless the latter has become the DESC SL WG). We may also want to include the Stars, Milky Way, and Local Volume SC, to see if they rely on photo-z to reject objects from their sample. One option would be to identify a point of contact that we can converse with, or another option would be to poll the general community. If we want to do a poll, perhaps questions like this would be useful.

1. Please rate how important LSST-provided photometric redshifts are to your science goals, e.g., on a scale of 1 to 5 (not important, insignificant, useful, important, essential).

2. Please rate how well the proposed LSST-provided photo-z meet your science needs, e.g., on a scale from 1 to 5 (not met, poorly met, adequate, well met, perfectly met).

3. Regardless of your answer to 2, please rate how often you might use the proposed LSST-provided photometric redshifts, e.g., on a scale of 1 to 5 (never, rarely, occasionally, regularly,

always).

2.3   Requirements Summary

The initial impression is that DM/QA might require a template-fitting estimator, and that machine-learning codes would be less favorable because they require a deep training set be made in advance (i.e., from the 10-year equivalent commissioning field).

3   Format of Photo-z in Level 2 Catalog

In this Section we discuss the format of the photo-z in the Level 2 catalog and options for compression. The photometric redshift products that the LSST DM team will provide in the Level 2 object catalogs are defined in DPDD, Table 4, "Level 2 Catalog Object Table" as:

- `stdColor` = "standard color", color of the object measured in "standard seeing", suitable for photo-z

- `stdColorErr` = uncertainty on stdColor

- `photoZ (float[2x100])` = photometric redshift likelihood samples, pairs of redshift and likelihood ($z, \log L$), computed using a published and widely accepted algorithm at the time of LSST

This desired format is set for an algorithm that returns $z, logL$, as opposed to a single "best" value and error, but as most estimators do return a posterior probability density function this is not a restrictive constraint. The posterior PDF will need to be compressed for storage within a `float[2x100]` array.

There are additional quantities that will be useful for Level 3 photo-z. The DESC photo-z WG has suggested that the DM-provided Level 2 DRP photo-z should be one that incorporates effective filter transmission curves for each individual epoch, because the required processing power for this is likely beyond the scope of any Level 3 photo-z estimator. However, Željko responded on the Jira thread that *"Although the Source table in DPDD doesn't list it explicitly (calibration docs do, but they are in state of flux right now), each Source must know (have a link to) its own transmission function (phi, see eq. 5 in the SRD). Once you know phi, and its standard version,*

phi_std, and you assume an SED, it's straightforward to compute photometric offset between the phi and phi_std systems. In practice, it will be somewhat more involved because we have a probabilistic photo-z, which implies a probabilistic, and not deterministic, SED.".

Recommendation: DM should update the DPDD to include the fact that each `Object.Source` **will have a link to its own unique filter transmission function (ϕ, Eq. 5 in the SRD); this is an important quantity for photo-z and the SC will want to know of it's existence and format.** And furthermore:

Recommendation: DM should establish a timeline for developing the algorithm (or releasing this algorithm, if it is already in place) that will be used to standardize the effective filter curves when generating the CoAdds, and verifying how well it meets specifications (e.g., systematics ~ 0.005 magnitudes).

3.1   Compression

Efficient $P(z)$ compression algorithms are in development. [6] present an alternative method for storing photo-z PDFs that requires less storage space. With their algorithm for sparse representation, "an entire PDF can be stored by using a 4-byte integer per basis function". As they describe, "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". The advantage in this method seems to be the accuracy, and the disadvantage is that the science user would need to convert the stored format into a $P(z)$ before they can use it (minor drawback). It is unclear to me how much additional processing is required for this conversion, and whether this is within DM's computational budget. The DESC photo-z WG has indicated that they expect compression techniques to be sufficient to store multiple algorithms' $P(z)$ results into the single database element of `Object.photoZ (float[2x100])`. This will increase both the advantages of compression and the required processing time. Also, Phil Marshall and Alex Malz are currently working on photo-z compression (2017-04-12).

3.2   Proposed Alterations

Two low-priority suggestions for minor changes to the photo-z content in the Level 2 database. (1) Many science users will simply want the photo-z and its uncertainty (e.g., the peak of the PDF and its 68% confidence interval). We recommend including this separately from the $P(z)$.

This could be a separate element in the database or e.g., the first element of `photoZ`. As these values are natural outcomes of most algorithms they should not increase DM processing. (2) If a template-fitting estimator is used, we need add a database element to store information about the best fit template. We will also have to make the templates publicly available. This information is important because it could help the user to evaluate whether the photo-z is sensible; e.g., allow them to weed out instances when an elliptical SED is best fit to a morphologically spiral object, or similar scenarios.

Recommendation: DM should expand the number of database elements allotted to photo-z from just the `float[2x100]` that will contain the $P(z)$ by adding one `float[4]` for a "best" photometric redshift and it's uncertainty (mean, median, 68%, 95%), and one e.g., `float[1]`, to encode the best-fit SED template (or synthesized mix of templates) when applicable. However, using 5 of the allocated 200 floats would probably be a suitable compromise.

4 Potential Options for Photo-z Evaluation

As an extremely preliminary demonstration of the types of tests a given photo-z estimator might undergo during the selection process to determine which will be included in the Level 2 database, we examine the performance of two off-the-shelf photo-z codes. We consider, for now, one template-fitting and one machine-learning estimator. In Section 4.1 we cover some existing work that was consulted in compiling these evaluation options. In Section 4.2 we describe the galaxy catalogs that we use. In Section 4.3 we describe the template-fitting photo-z estimator BPZ and the machine-learning code TPZ. Section 4.4 demonstrates some analysis techniques for comparison of different estimators' photo-z results. We emphasize that the parameters of the estimators themselves have not been optimized to return the most accurate photo-z, and some minor mistakes have been made in these initial runs, but that this is fine as we are simply demonstrating an evaluation process. In Section 4.5 we write down some ideas for future tests of the photo-z estimators that would evaluate their performance in real world situations, and that are similar to the types of early science that might be done with LSST photo-z.

4.1 Relevant Existing Work

DESC photo-z WG – This science community is full engaged in the development of photo-z routines and their optimization for LSST. Here we list just other relevant existing work.

Relevant photo-z testing papers – [9] 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 – [14] 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. As I understand it, they first use an empirical technique with a large training set to estimate the redshift and it's 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 – Lucy Halperin used what we call "Brown" catalog[2] 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. We 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 our use of the "Brown"-based galaxy catalog. We do not use this catalog in Section 4.

---

[2]We call it the Brown catalog because it uses the SEDs from [3]

4.2 Simulated Catalogs

We use a randomly chosen 30000 galaxy test subset of the LC_DEEP_GONZALEZ2014A cata-log, which is based on the Millennium simulation [15] and the galaxy formation models of [8] and constructed using the lightcone techniques described by [13]. 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 require-ment 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 algo-rithms' 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 simu-late 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 sur-vey 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 algo-rithm 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. 11). Assuming that the LSST will cover a spectro-scopic 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 photo-metric 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.

4.3 Example Photo-z Estimators

Bayesian Photometric Redshifts (BPZ; Benítez 2) is a template-fitting algorithm with a mag-nitude prior[3]. We use all default parameters, including the i-band for the magnitude prior,

---

[3]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 [10], 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_{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.

Nearest-Neighbors in Color (NNC; Graham et al. in prep) 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.

4.4 Techniques for Analysis and Comparison of Photo-z Estimators

In this section we demonstrate several analysis techniques for comparing the output of different photo-z estimators, discussing each in turn below and then listing some potential future techniques in Section 4.5.

The z_{true}–z_{phot} Diagram

Figure 1 shows the photo-z results in z_{true}–z_{phot} diagrams, which are a typical way to visually asses 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_{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 $z_{\mathrm{phot}} > 1.5$. 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). [8] 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 it's 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, $\Delta z_{(1+z)} = (z_{\mathrm{spec}} - z_{\mathrm{phot}})/(1 + z_{\mathrm{phot}}$. 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 $\overline{\Delta z_{(1+z)}}$; and the fraction of outliers, f_{out}, which is the fraction of galaxies with $|\Delta z_{(1+z)}| > 0.06$ and $> 3\sigma_{\mathrm{IQR}}$ (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
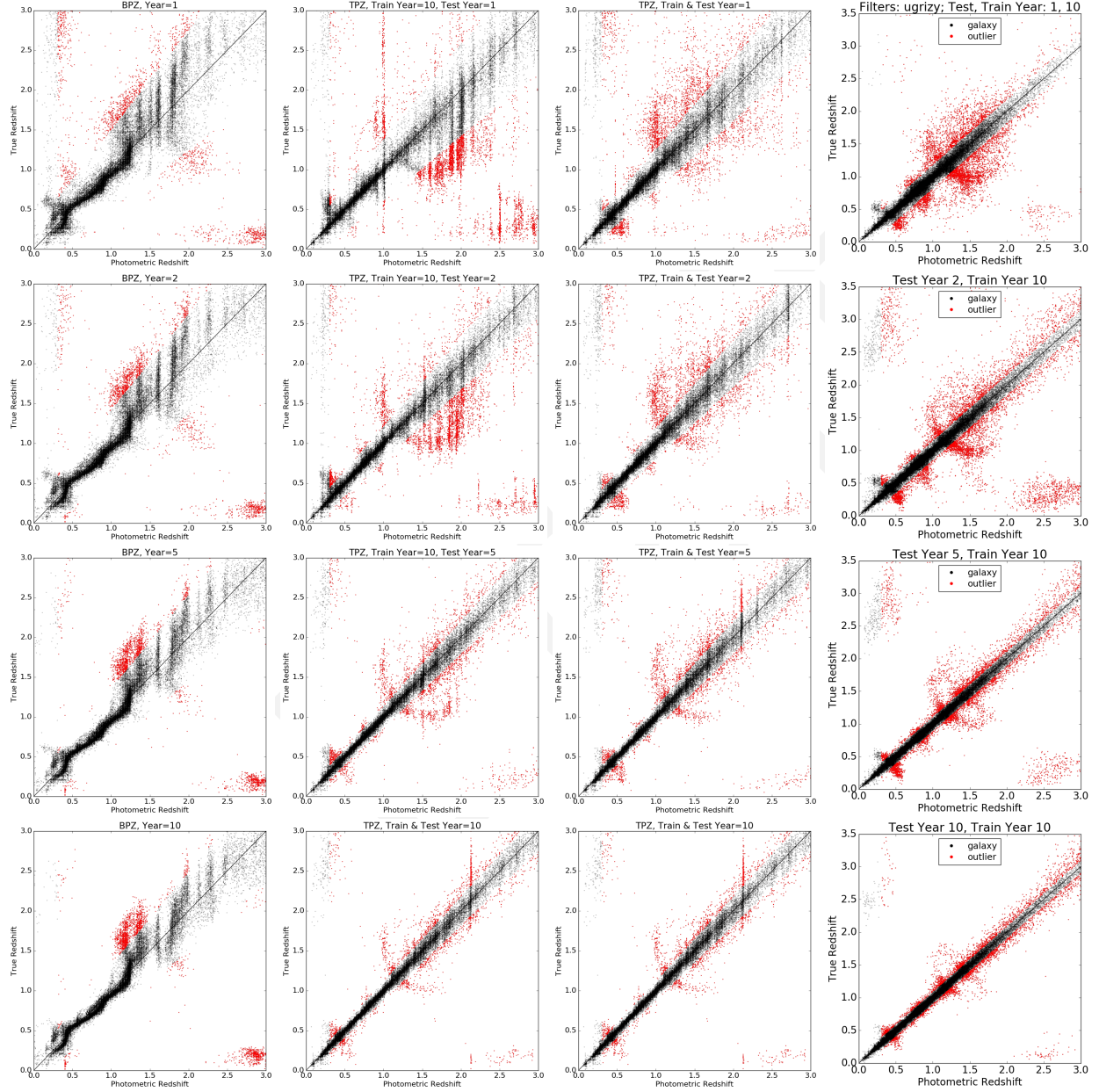
FIGURE 1: Examples of z_{true}–z_{phot} plots for a variety of algorithms (by column), for 1, 2, 5, or 10 years of survey time elapsed (top to bottom). Galaxies that are statistical outliers are shown in red. **Left:** results for the BPZ algorithm. **Center-left:** results for the TPZ algorithm with 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.

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., $0.3 \leq z_{\mathrm{phot}} \leq 3.0$) 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, δz_{phot}

In Figure 4 we demonstrate a way to assess the photo-z uncertainties, δz_{phot}, that come out of the estimators: we plot z_{phot}–δz_{phot} in the main axis, and above and to the side plot the distributions in δz_{phot}, z_{phot}, and for comparison, z_{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 δz_{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

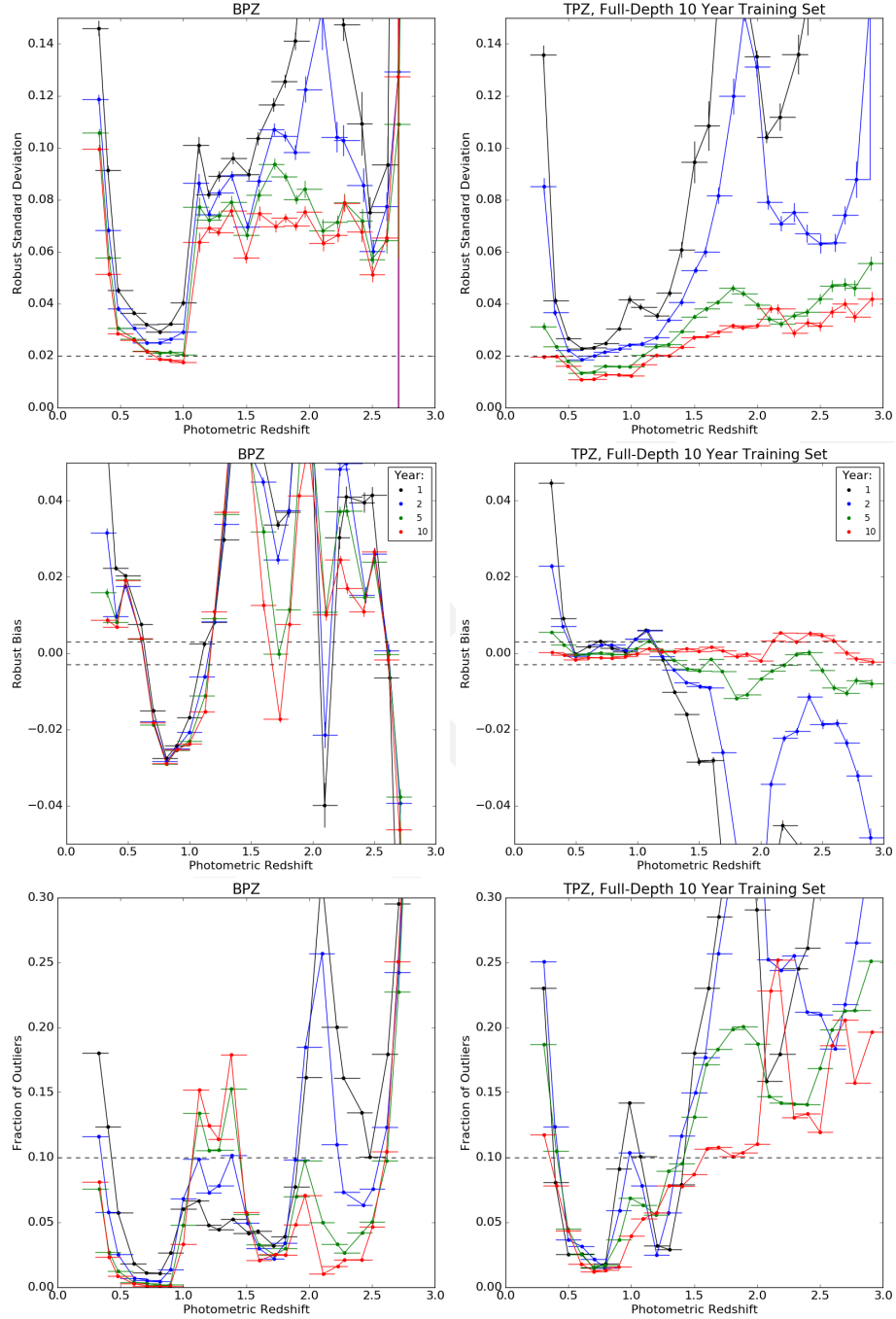In Figure 6 we show an example of a quantile-quantile plot using the true vs. the photometric

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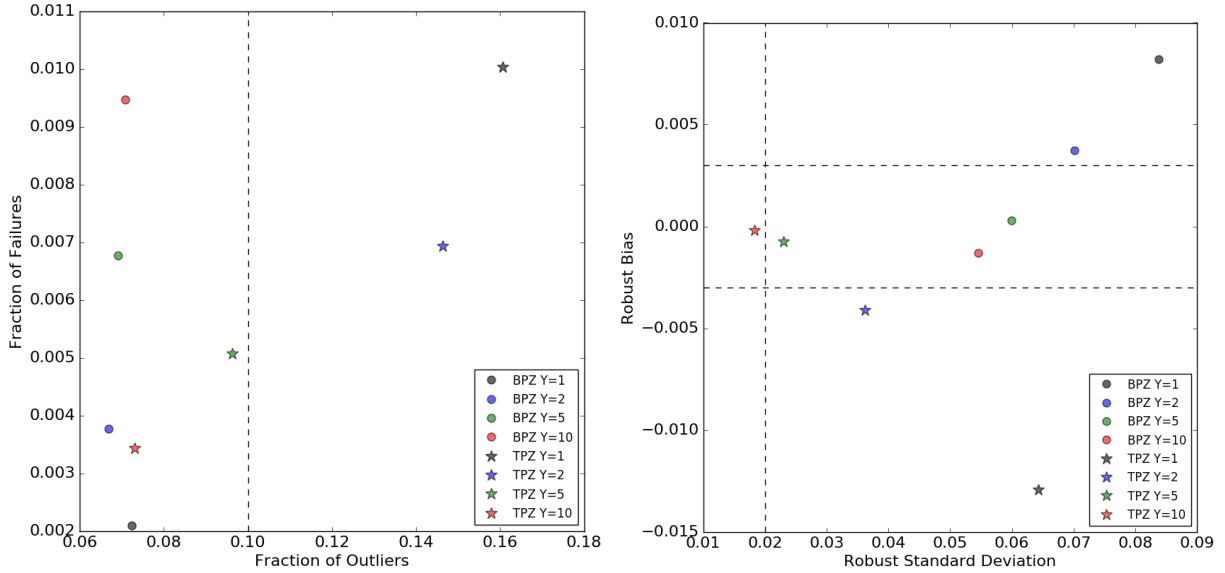
FIGURE 3: Examples of how to compare statistical measures over $0.3 \leq z_{\mathrm{phot}} \leq 3.0$ from different photo-z estimators by plotting one against the other: fraction of failures and outliers (left), and the robust bias and standard deviation (right). In this case we're comparing the statistical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

redshift. Each point represents the z_{true} and z_{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_{phot} would match that of z_{true}. In Figure 6 we can see for both BPZ and TPZ that this is not the case, but that BPZ is worse.

4.5   Ideas for "Real-World" Science Tests

Instead of comparing the general performance of photo-z estimators, we could compare them in specific real-world situations that represent some of the early science uses of Level 2 photo-z. Here is a short list of ideas.

Cluster Identification: take a line-of-sight from the simulated catalog with a known cluster, run the photometry through different algorithms, see which ones identify the cluster first/best.
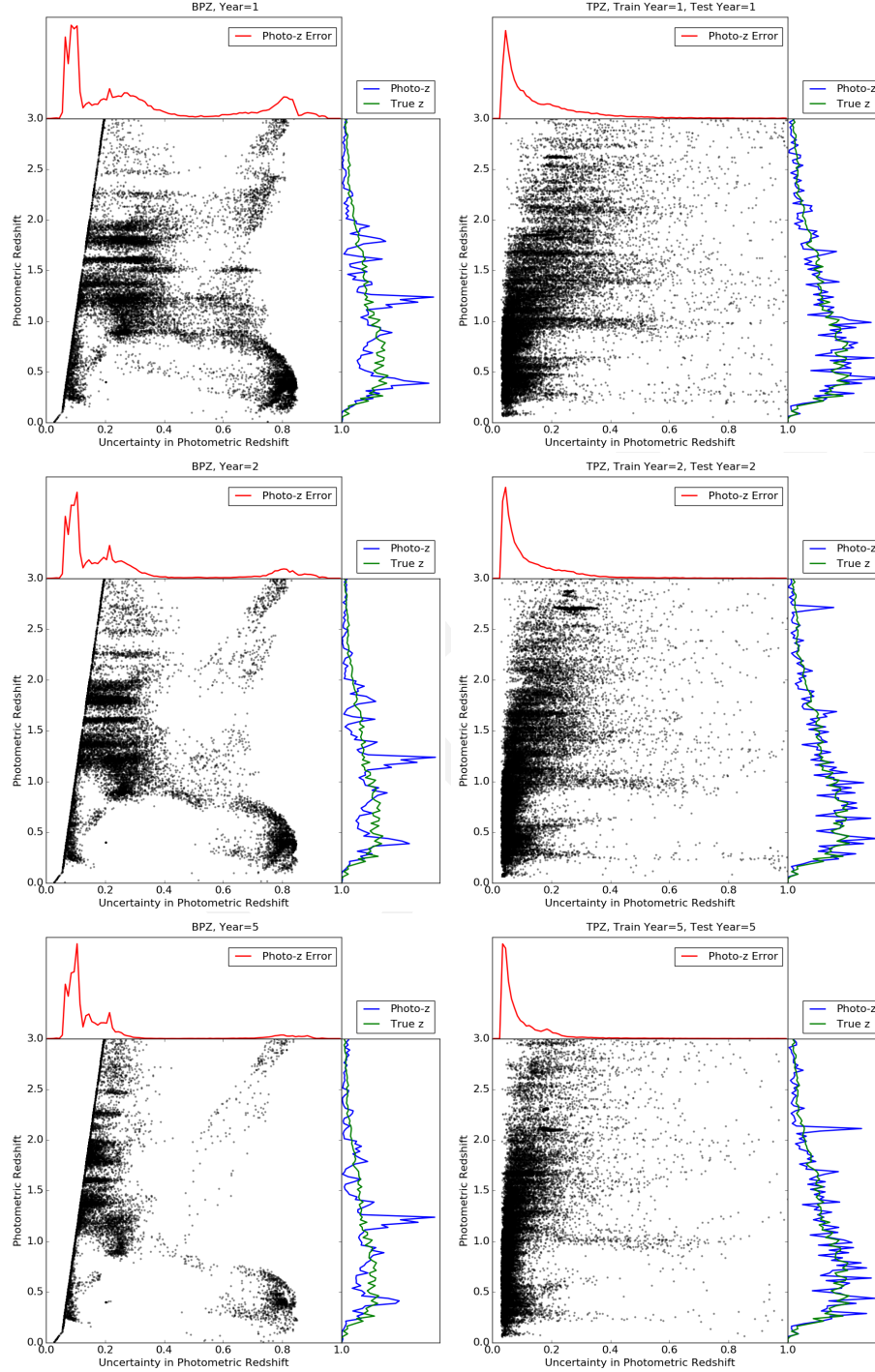
FIGURE 4: Examples of plot to compare the photo-z uncertainty (δz_{phot}) between algorithms with z_{phot}–δz_{phot} plots from the BPZ (left) and TPZ (right) estimators for simulated catalogs with photometric uncertainties at 1, 2, and 5 years of LSST (top to bottom). Red lines show the distribution of photo-z errors; blue and green lines compare the distributions of true and photometric redshifts.
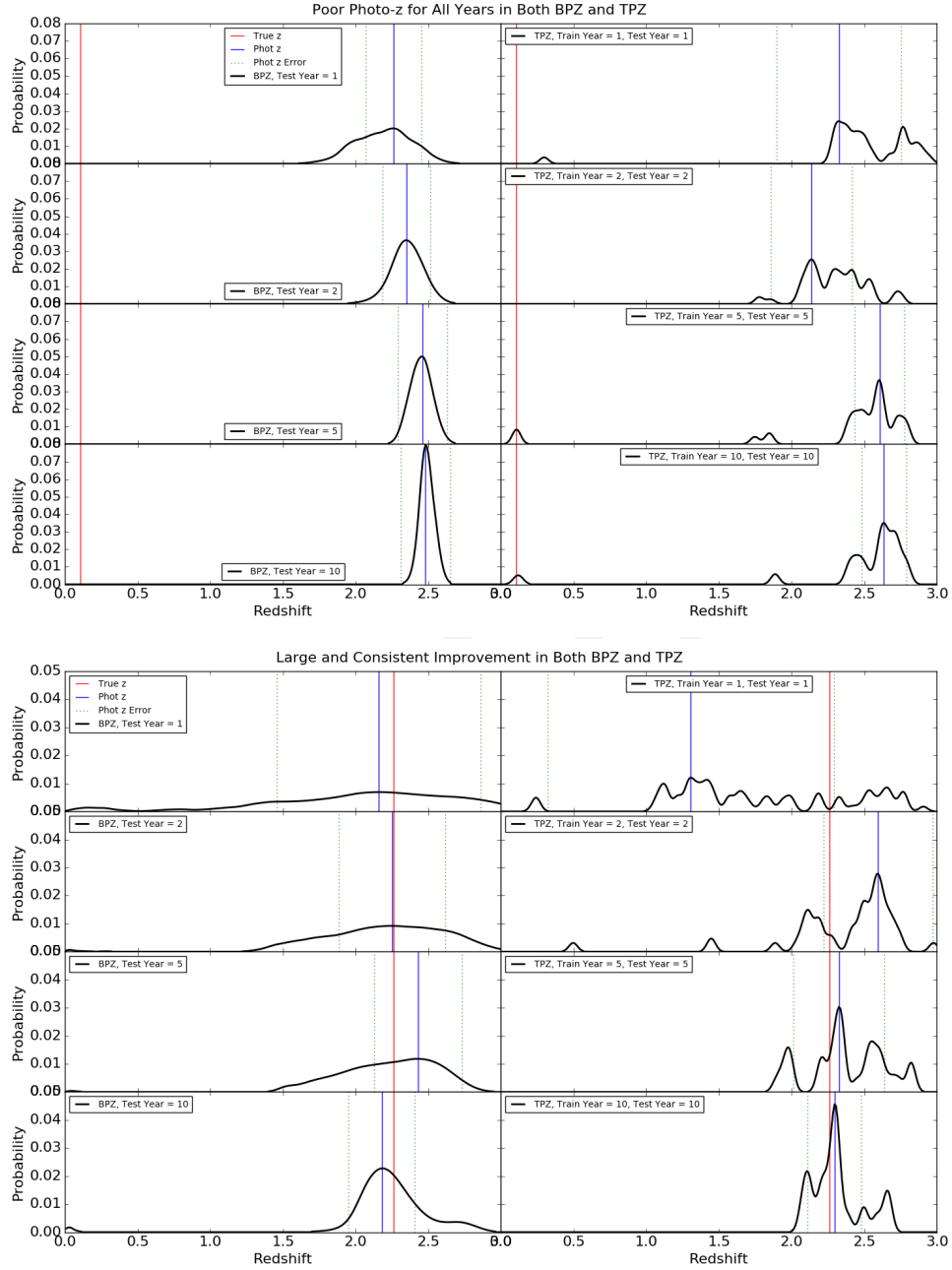
FIGURE 5: Examples of the posterior probability density functions for two test galaxies in all of our simulations: BPZ (left) and TPZ (right) for photometric uncertainties like 1, 2, 5, and 10 years of LSST (rows from top to bottom). In the top panel we choose a galaxy that return inaccurate and imprecise 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.
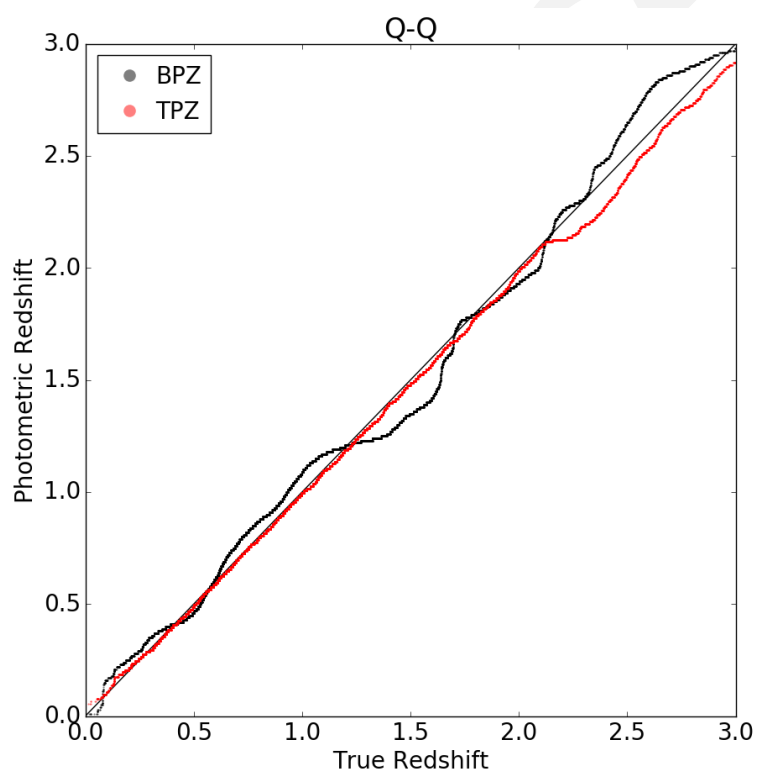
FIGURE 6: Example of a Q-Q plot, using z_{true} and z_{phot}, from the BPZ and TPZ estimators.

Supernova Classification: Rahul is working on how including photo-z helps SN classification, so perhaps we could see which algorithm produces the most useful photo-z for likely SN hosts.

Star/Galaxy Photometric Separation: The Euclid catalog doesn't include $z = 0$ stars, but we could insert some and see how well the different algorithms are at picking them out, either by including stellar templates to actively identify them, or by passively identifying stars by how they fail to return a decent photo-z.

5 Conclusions

For now, our interim recommendations are listed up front in the Introduction.

References

[1] Beck, R., Dobos, L., Budavári, T., Szalay, A.S., Csabai, I., 2016, MNRAS, 460, 1371, ADS Link

[2] Benítez, N., 2000, ApJ, 536, 571, ADS Link

[3] Brown, M.J.I., Moustakas, J., Smith, J.D.T., et al., 2014, ApJS, 212, 18, ADS Link

[4] Carrasco Kind, M., Brunner, R., 2013, TPZ: Trees for Photo-Z, Astrophysics Source Code Library

[5] Carrasco Kind, M., Brunner, R.J., 2013, MNRAS, 432, 1483, ADS Link

[6] Carrasco Kind, M., Brunner, R.J., 2014, MNRAS, 441, 3550, ADS Link

[7] Dahlen, T., Mobasher, B., Faber, S.M., et al., 2013, ApJ, 775, 93, ADS Link

[8] Gonzalez-Perez, V., Lacey, C.G., Baugh, C.M., et al., 2014, MNRAS, 439, 264, ADS Link

[9] Hildebrandt, H., Arnouts, S., Capak, P., et al., 2010, A&A, 523, A31, ADS Link

[10] Ilbert, O., Arnouts, S., McCracken, H.J., et al., 2006, A&A, 457, 841, ADS Link

[11] Le Fèvre, O., Tasca, L.A.M., Cassata, P., et al., 2015, A&A, 576, A79, ADS Link

[12] Leistedt, B., Hogg, D.W., 2016, ArXiv e-prints, ADS Link

[13] Merson, A.I., Baugh, C.M., Helly, J.C., et al., 2013, MNRAS, 429, 556, ADS Link

[14] Sánchez, C., Carrasco Kind, M., Lin, H., et al., 2014, MNRAS, 445, 1482, ADS Link

[15] Springel, V., White, S.D.M., Jenkins, A., et al., 2005, Nature, 435, 629, ADS Link