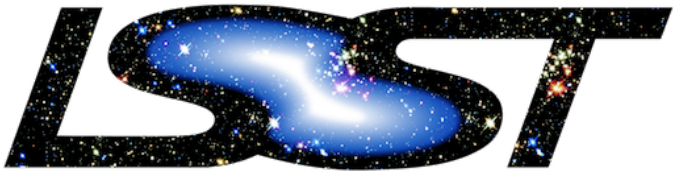

LARGE SYNOPTIC SURVEY TELESCOPE

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, and the DM System Science Team.

DMTN-049

Latest Revision: 2019-12-16

D R A F T

Abstract

This document discusses options for the generation and validation of photometric redshifts for the LSST Data Release `Objects` catalog. This is a *living document* which will progress over time as the photo-z attributes are defined, one or more estimators are selected, and the validation process is established. Contributions from the science community are solicited regarding the contents of this document.

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

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 (e.g., apparent magnitudes in multiple filters) instead of on its spectral features (e.g., emission and absorption lines). Redshift (distance) is a key component of many science goals that will be pursued with the LSST data. Since it will be impossible to obtain spectra for the billions of galaxies that LSST will observe, photometric redshift estimates will be necessary.

Typically, photometric redshift estimators 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 (e.g., point estimates, full posterior probability density functions, PDFs, or redshift distributions in tomographic bins). For this reason, several research groups in the science community are already planning to generate multiple kinds of photo-z (e.g., the Dark Energy Science Collaboration).

However, it would be scientifically prohibitive if *all* LSST users had to generate their own photo-z estimates, as this is a computational intensive calculation. Furthermore, it is a requirement that the LSST Data Management System (DMS) calculate photo-z and store them in the `Object` catalog. 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 estimator(s) would be installed by DM into the DR pipeline and run at scale, and/or a user-generated photo-z catalog could be ingested and federated with the `Object` catalog. DMS-REQ-0046

The purpose of this document is to clarify how the DMS will fulfill the requirement to generate and serve `Object` photo-z with each data release during Operations; to define the attributes and performance of this data product, as well as its associated documentation; and to describe how contributions from the LSST science community, which has a considerable wealth of expertise in generating photo-z catalogs, will be incorporated.

2 Proposed Timeline

The following is a proposed path to prepare the DMS to generate and serve `Object` catalog photo-z as part of the data release processing pipeline during Operations.

2.1 Q2 2020: Define Minimum Attributes

A preliminary proposal for the minimum attributes (e.g., the type, outputs, performance) of the `Object` catalog photo-z is put forth in § 3. Community input on the definition of these attributes will be solicited via the science collaborations and through interaction at meetings. This iterative process should end by Q2 2020, when the LSST Project will update to this document with the definition of photo-z attributes.

2.2 Q4 2020: Evaluate Algorithm(s)

A preliminary proposal for the process by which potential photo-z algorithms should be evaluated is put forth in § 4. Community input on the evaluation criteria and the benefits/drawbacks of potential algorithms will be solicited via the science collaborations and through interaction at meetings. This iterative process should end by Q4 2020, when the LSST Project chooses one or more algorithms to be implemented, and updates this document with the details of the selection.

2.3 Q4 2022: DMS Implementation and Validation

A preliminary proposal for the photo-z implementation process and validation tests are assembled in § 5 and 6, respectively. The DM team is responsible for implementing and validating the chosen algorithm(s) into the data release pipeline, along with any needed supporting data sets (e.g., templates, spec-z catalogs). Lists of data products associated with photo-z are collected in § 7. Community assistance with the process of implementing and validating the photo-z algorithm will be solicited via the science collaborations and through interaction at meetings, and will include access to commissioning data previews. This process should end by Q4 2022, i.e., the end of LSST Construction and the start of Operations.

2.4 Data Releases

During LSST Operations, the DMS would produce and validate photo-z for the `Object` catalog, and these photo-z are available at the time of all data releases, along with any and all supporting materials such as documentation or spectral templates. The LSST Operations project may solicit and collect community feedback on the photo-z, and return to any stage of this path and change the attributes, algorithm, implementation, and/or validation of `Object` catalog photo-z for future data releases.

2.5 Option: Federate a Community Photo-z Catalog

As described in § 3, the DMS as delivered to the Operations Project will provide a minimal scientific capability with respect to `Object` catalog photo-z. If that is rendered obsolete by community efforts, then the superior product should be ingested and federated.

To facilitate this option might require providing the community team(s) with access to data release previews[1] so that they may train and calibrate their photo-z algorithm, and minimize the time between data release and federation of a new photo-z catalog to the `Objects` table.

Options to federation and integration into the `Objects` catalog might include the LSST Operations agreeing to host a "photo-z server" within the LSST Science Platform. For example, a system like the Dark Energy Survey's Science Portal, an infrastructure for organizing input catalogs, installing photo-z algorithms, training and running them, and evaluating their output[2], as described by Gschwend et al. [10]).

2.5.1 The Science Impacts of a Data Release Without `Object` Photo-z

The option to federate a community-generated photo-z catalog leads to this question: *If there will likely be a superior community photo-z anyway, should the LSST project avoid installing a photo-z algorithm in the DMS, and instead simply wait for the community to generate a catalog?* There are several significant risks and drawbacks to this option.

[1] Data release previews are envisioned to be some small, e.g., $\sim 10\%$, amount of a data release made available early, e.g., weeks to months, before the full release to enable community feedback and preparations prior to the full release.

[2] A series of YouTube tutorials about the DES Science Portal are available at `https://www.youtube.com/playlist?list=PLGFEWqwqBauBIYa8H6KnZ4d-5ytM59vG2`.

- An `Object` table without photo-z at the time of data release is a problem for brokers, unless alerts are instead (or additionally) associated to an older DR's `Object` catalog that has photo-z. Brokers require host-galaxy photo-z to optimally classify and prioritize transients for follow-up, and plan to obtain this information via each alert's associations with nearby `Objects` from *the most recent DR*.
- This option does not satisfy a literal interpretation of the requirement that the *"DMS shall compute"* photo-z. DMS-REQ-0046
- The `Object` photo-z might be tailored to the specific science case of the community team and might not serve the broader science use-cases.
- There is no initiative or reward for the community team which has generated the photo-z, except perhaps citations to their photo-z catalog.
- There would be no `Object` photo-z if no community team generates and donates a catalog, which is a risk for the science use-cases described in § 8.2

3 Proposed Minimum Attributes

These proposed minimum attributes will be refined via an iterative process between the LSST project and the broader scientific community.

The following attributes are defined based on the guiding principle that the DMS-generated `Object` photo-z should, at minimum, *meet the basic science needs for communities which will not or cannot generate custom photo-z*, and also DM's internal use-cases. The basic science needs and use-cases for `Object` photo-z which were used to propose this set of preliminary minimum attributes are described in § 8.

A *basic* science need would not include, for example, photo-z of the quality required for major cosmological advances. The development of such photo-z algorithms is an active research topic within the Dark Energy Science Collaboration [20], and a significant effort which the LSST Project should not (and could not) attempt to replicate.

Type of Algorithm

The `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 § 7.4). The main motivation for this is galaxies-related science (§ 8.2.3) and internal (§ 8.1) use-cases, which would use the best-fit template to derive additional galaxy properties.

Output Format

The `Object` catalog photo-z should include the posterior distribution function and a point estimate with an uncertainty, as well as reliable uncertainties and/or flags to help the novice user avoid mis-applying the results, or over-estimating their significance. For template-fit photo-z results, an identifier for the best-fit template should be provided, and those templates be made publicly accessible. A list of the photo-z outputs (and documentation) is put forth in § 7.3.

Performance

The `Object` catalog photo-z could have a point-estimate accuracy of $\sim 10\%$ and still meet the basic science needs. The photo-z results should have a standard deviation in $z_{\rm true} - z_{\rm phot}$ of $\sigma_z < 0.05(1 + z_{\rm phot})$, and a catastrophic outlier fraction of $f_{\rm outlier} < 10\%$, over a redshift range of $0.0 < z_{\rm phot} < 2.0$ for galaxies with $i < 25$ mag galaxies.

4 Proposed Evaluation Criteria

These proposed evaluation criteria will be refined via an iterative process between the LSST project and the broader scientific community.

These criteria may also apply to community-generated catalogs proposing to be federated with the `Objects` table (or be otherwise served at scale to the broader science community).

Minimum Attributes
The algorithm should meet the minimum attributes defined in § 3, and serve both the basic science needs of the science community and the internal use-cases of the LSST project (§ 8). Performance beyond the basic needs, and the ability to provide higher-quality photo-z results, should obviously also be considered favorably. Evaluating and comparing the performance of photo-z estimators is discussed further in § 4.1. The proposed minimum performance is likely achievable for LSST data with current photo-z estimators, as demonstrated by, e.g., Graham et al. [9], Schmidt et al. (2020, in prep.).

Scientific Utility
The selected algorithm(s) should serve as wide a variety of science needs as possible. Demonstrated success with other wide-field optical surveys, previous application to surveys that overlap with LSST (for results comparisons), and general community support or adoption of the estimator should all be considered.

User Experience
The photo-z estimator's algorithms and output should be straightforward to understand and easy to access. The selected algorithm(s) should have detailed, publicly accessible documentation (e.g., a journal article, a website, a schema) to facilitate community use. Desired photo-z outputs, access methods, and documentation are discussed in § 7.3.

Computational Feasibility
The selected algorithm should take as inputs quantities that the LSST pipelines are planning to produce (§ 7.1). The necessary data – from LSST or elsewhere – for training and/or calibration must exist by the time of DR1 (e.g., galaxy catalogs with spectroscopic redshifts and LSST photometry; § 7.2). The computational needs of the selected algorithm must fit within the available resources of the LSST data facility.

Personnel Requirements
New algorithmic development is beyond the scope of DM, and the selected algorithm should not need any expansion or maturation prior to installation in the DMS.

4.1 Evaluating Photo-z Estimators' Performance

Criteria for evaluating the statistical performance of different photo-z estimators during the selection process overlap with the validation tests discussed in § 6, especially the truth comparisons.

Examples of Photo-z Estimator Comparisons

- Hildebrandt et al. [11] tested 18 different photo-z codes on the same sets of simulated and real data and found no significantly outstanding method.
- Dahlen et al. [7] tested 11 different photo-z codes on the CANDLES data set (U-band through infrared photometry) and also find that no method stands out as the "best", and that most of the photo-z codes underestimate their redshift errors.
- Sánchez et al. [17] used 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 [2]) 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."
- Tanaka et al. [19] provides a comparative analysis of several photo-z algorithms applied to their data set from the HSC Strategic Program. Their website[3] provides comparative analysis plots for each of them.
- Schmidt et al. (2020, in prep.) statistically compare the posterior probability distribution functions produced by 12 photo-z estimators for a mock data set that is representative of the LSST, identifying some biases and shortfalls in both the produced PDFs and the evaluation methods used to analyze them.

[3] https://hsc-release.mtk.nao.ac.jp/doc/index.php/photometric-redshifts/

5 Proposed Implementation Process

This proposed outline for the implementation of selected photo-z estimator(s), and the roles of the commissioning data and community contributions, will be refined via an iterative process between the LSST project and the broader scientific community.

In Q1 2021, the evaluation of different photo-z estimators will have concluded and the implementation of the selected estimator(s) into the data release pipeline will begin (§ 2). It is expected that the implementation and validation of the selected the photo-z estimator(s) will take up to Q4 2022, i.e., until the end of LSST Construction and the start of Operations.

The basic steps to take place during implementation would be as follows:

- prepare the LSST data inputs to the photo-z estimators (§ 7.1)
- prepare the training and calibration data (§ 7.2)
- install the needed codes (a technical aspect to be described elsewhere)
- run the estimator(s) (on its own and/or embedded in the data release pipeline)
- validate the photo-z estimator outputs (§ 6)
- store the results in the `Object` catalog (§ 7.3 and 7.4)
- ensure output schema and access methods are documented (§ 7.3)

The Role of Commissioning Data During Implementation

The photo-z implementation process will rely on commissioning data. In **late 2021** the first phase of commissioning, Data Preview 1 (DP1) with LSST ComCam, would begin and include, e.g., tests of the scheduler; tests of image quality, depth, astrometry, and photometry; and a 20-year depth test to stack images over a range of conditions [LSE-79]. DP1 will be useful as preliminary test data during the implementation of the photo-z estimators, and help to, e..g, test the formats of the input and output data, create visualizations for the validation codes, etc. In **mid 2022** the second phase of commissioning, Data Preview 2 (DP2) with the LSST Science Camera, would begin and include wide area surveys to the full 10-year depth in addition to a 20-year depth stack and tests of image quality, etc., as in DP1 [LSE-79]. DP2 will provide the data set for the photo-z validation process described in § 6, and also some of the needed training data for LSST photo-z described in § 7.2.

Infrastructure for Community Contributions to Implementation

It is desirable that the photo-z implementation process be open to contributions from the broader community. To enable this, the community will need access to LSST-like data products that are generated from, e.g., HSC survey data or LSST commissioning data from the LSST Science Camera (i.e., data release preview 2). The community would also benefit from having this access through the LSST Science Platform, to facilitate the communal development and sharing of codes related to validation. For example, a system like the Dark Energy Survey's Science Portal, though which users may train, run, and evaluate photo-z algorithms Gschwend et al. [10].

6 Proposed Validation Tests

These proposed validation tests will be refined via an iterative process between the LSST project and the broader scientific community.

Regardless of whether the LSST photo-z are a DMS-generated data product or a federated community-generated catalog, validation tests and quality assessment diagnostics will be necessary to ensure the results meet performance expectations (§ 3) and to facilitate science applications among the broader LSST user community.

Below is a compilation of potential validation tests, some of which overlap with the evaluation criteria described above (especially the truth comparisons). These lists are based in part on a brainstorming session during the LSST Project and Community Workshop's session on Photometric Redshifts on Aug 14 2019[4].

Journal articles that demonstrate validation processes for photo-z from multi-band wide-area surveys include *"DES science portal: Computing photometric redshifts"* [10]; *"Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1"* [19]; and *"On the realistic validation of photometric redshifts"* [1]. Furthermore, DESC is currently investigating methods of validating accuracy of probability distributions from a photo-z algorithm (see, e.g., Schmidt et al. 2020, in prep).

Truth Comparisons

Catalogs of true redshifts can be obtained by withholding some fraction of the training set or by cross-matching to external spectroscopic catalogs. Validation metrics should include at least those used to define the minimum performance of the LSST photo-z for basic science needs (§ 3). A list of potential metrics might include the following, and targets or limits on these metric values might apply to subsets in magnitude, color, or redshift:

- plots of $z_{\rm true}$ *vs.* $z_{\rm phot}$ for visual inspection
- standard deviation and bias in $\Delta z = (z_{\rm true} - z_{\rm phot})/(1 + z_{\rm true})$
- fraction of (catastrophic) outliers, e.g., $\Delta z > 3\sigma$ or > 0.06 (> 1)
- quantile-quantile (q-q) plots evaluate the shape of $P(z)$

[4] E.g., slide 14 of https://docs.google.com/presentation/d/1GEahvDQXIjSL4lLVjDlZHV5zpXLhGQfHwVtucs72Ajg/edit?usp=sharing

- probability integrated transform, $PIT(z_{\rm phot}) = \int_0^{z_{\rm phot}} P(z)\,dz$, calculated for all test galaxies, should be flat for well-estimated $P(z)$ [15]
- the continuous ranking probability score, CRPS, should be a lower value for better estimated $P(z)$, as described in [15]
- redshift confidence for point estimates, $C(z_{\rm phot}) = \int_{z_{\rm phot}-0.03}^{z_{\rm phot}+0.03} P(z)\,dz$
- loss and risk parameters that characterize the photo-z estimates:
  $L(\Delta z) = 1 - \left(1 + \left(\frac{\Delta z}{\gamma}\right)^2\right)^{-1}$, where γ is a characteristic threshold (e.g., 0.15), and:
  $R(z_{\rm phot}) = \int P(z)\,L(z_{\rm phot}, z)\,dz$, as described in Tanaka et al. [19]
- the conditional density estimation (CDE) loss as described in Section 4.2 of Schmidt et al. (2020, in prep). (Note that the CDE does not strictly require the true posterior be known.)

Sanity Checks

These are tests that can be done using all LSST `Objects` and do not require a truth catalog.

- the uncertainty in $z_{\rm phot}$ should correlate with photometric error
- the star/galaxy flag parameter should agree with photo-z (stars have $z_{\rm phot} = 0$)
- evolution of $N(z)$ to higher-z for samples with fainter magnitudes, redder colors

Scientific Applications

The following might not be included in the formal validation process, but represent analyses that the broader scientific community may want to do to inform their use of the LSST photo-z.

- 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
- tomographic bin analysis, as in [6]

7 Data Products Related to LSST Photo-z

The contents of the following sections will be refined over time to eventually describe the data products input and output from the selected photo-z estimator(s), and contributions from the scientific community are solicited for each of the following topics.

This section is a preliminary collection of details related to generating and serving LSST photo-z, such as the LSST data products that will be needed as input to photo-z estimators (§ 7.1), the required training and calibration data from LSST and external sources (§ 7.2), the outputs' format, access methods, and documentation (§ 7.3), and options for compressing photo-z results to potentially store the results of multiple estimators (§ 7.4).

7.1 Inputs to Photo-z Estimators

It is important to ensure that all measured quantities needed by photometric redshift estimators are going to be computed and included in the `Object` table.

Aside from the fluxes and/or apparent magnitudes and errors for each LSST filter, which will be provided in the `Object` catalog, the color properties in the `Object` table might be used for photo-z. *"Colors of the object in 'standard seeing' (for example, the third quartile expected survey seeing in the i band, ~0.9 arcsec) will be measured. These colors are guaranteed to be seeing-insensitive, suitable for estimation of photometric redshifts"* [LSE-163]. In the `Object` table the relevant elements are:

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

Additionally, measured quantities such as the galaxy size, shape, radial profile, 'clumpiness', or surface brightness; the DCR correction (or residual); or a parameter that represents the clustering density within some radius (e.g., 2 Mpc) might all be useful (e.g., as priors) for photo-z algorithms. The effective transmission function (ϕ; Eq. 5 in LPM-17), which will be provided for all `Sources` either in the catalog or as a link, is another useful quantity for photo-z estimators.

7.2 Training and Calibration Data

Regardless of whether the `Object` catalog photo-z are generated by the DMS or an ingested community catalog, some training and calibration data from LSST will be needed.

Spectroscopic Redshifts
Deep multi-band LSST photometry for spectroscopic fields like COSMOS, and/or WFD-depth LSST photometry that overlaps multiplexed spectroscopic surveys like DESI/4MOST, which is obtained either during commissioning or early in Operations year 1, will likely be necessary to produce photo-z for DR1.

Wide Area Imaging
Some photo-z methods have requirements other than spec-z fields: e.g., Sánchez & Bernstein [16] 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).

Data Previews
For the community to participate in the training or calibration of a photo-z algorithm prior to a data release, it will probably be necessary to release a small ($\sim 10\%$) but representative "DR preview" in advance of each data release during Operations.

7.3 Output Schema, Access Methods, and Documentation

This section gathers details regarding what the photo-z outputs should be, how they should be accessed, and how they should be documented.

Output Schema
The format of the photo-z output is one of the minimum attributes that must still be defined (§ 3). The photo-z outputs must provide all the necessary inputs to the validation tests, to be defined in § 6. The currently proposed `Object` catalog table elements related to photo-z are defined in LSE-163 and provided in § 9.4, and summarized here:

- posterior probability distribution function (likelihood over redshift)
- a single point estimate with an uncertainty

- quantities related to the posterior (e.g., mode, mean, skewnewss)
- flags (e.g., potential catastrophic outlier, failure mode, consistent with z=0)

Access Methods

The user experience is one of the proposed selection criteria for the LSST photo-z algorithm (§ 4). Some examples of publicly released photo-z catalogs which were prepared with a user experience that might be desirable for the LSST photo-z include the Dark Energy Survey's Science Portal to serve photometric redshifts [10] and the Hyper SuprimeCam Subaru Strategic Program [19][5].

If the LSST photo-z are not made available in either the `Objects` table or in a federated or joinable catalog – for example in the case where a community-generated photo-z catalog is replacing the DMS-generated catalog (§ 2.5) – and are instead made available via, e.g., a "photo-z server" (as in [10]), then at least the `Object` catalog ID of the most recent data release should be a queryable parameter.

If the results of multiple algorithms are generated, compressed, and stored in the `Objects` table, then decompression should be straightforward for the user (§ 7.4).

Documentation

Appropriate types of documentation might include published journal articles, GitHub repositories, websites, or other online documentation resources (e.g., `https://readthedocs.org/`). Whatever the format, the documentation contents should include:

- general description of the estimator
- adaptations made to ingest LSST data (compared to past applications)
- an analysis of the training, calibration, and validation processes
- a full list of all inputs and outputs (i.e., a schema browser)

7.4 Storage and Compression

Regardless of whether the `Object` catalog photo-z are generated by the DMS or an ingested community catalog, the stored values are subject to the storage space allotted in the `Objects`

[5] `https://hsc-release.mtk.nao.ac.jp/doc/index.php/photometric-redshifts/`

table as described in § 9.4. However, both the posteriors and point estimates from several different photo-z estimators could be compressed and stored in this allotted space. Furthermore, given the variety of use-cases and the fact that different photo-z estimators produce different results (Schmidt et al. 2020, in prep.), the option to compute, compress, and store estimates from multiple algorithms in the 2×95 float might be scientifically desirable.

Efficient $P(z)$ compression algorithms are in development, such as Carrasco Kind & Brunner [3] and Malz et al. [14]. Carrasco Kind & Brunner [3] 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. [14] 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, and decompression must be fast and easy for users.

8 Example Use-Cases for LSST Photo-z

This section contains an incomplete, non-exhaustive summary of a variety of internal and scientific use-cases for the LSST `Object` catalog photo-z. These use-cases inform the minimum attributes and selection criteria proposed in § 3 and 4.

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[6].

8.1 Internal DMS Use-Cases

DM's galaxy photometry outputs are being developed with the goal of feeding photometric redshift algorithms, so the computation of photometric redshifts is likely to be a part of the science validation process for LSST photometry. 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 such a simple SED-fit photo-z could also serve the scientific use-cases is undetermined, because the photometric validation process is not yet defined or written. The internal use-case described in § 9.2 – of needing photo-z in order to assess catalog completeness for low- and high-redshift `Objects` – is also likely to be served by a simple SED-fit photo-z estimate.

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.

[6] Thanks to Sam Schmidt, Chris Morrison, Sugata Kaviraj, Gautham Narayan, Lauren Corlies, Travis Rector, Tina Peters, Alex Malz, Dara Norman, Stephen Smartt, and other participants from the science community.

8.2 Scientific Use-Cases

A variety of potential scientific applications for the `Object` photo-z are discussed in turn. These scientific use-cases should be used to inform the minimum attributes and selection criteria proposed in § 3 and 4. A summary of the commonalities between science use-cases for photo-z is provided in § 8.2.8.

8.2.1 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, and the photo-z results are subject to established science requirements for dark energy cosmology [20]. For example, weak lensing and large scale structure require ensemble measurements of $N(z)$ and thus require a full posterior PDF, 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 (DESC) is developing specialized photo-z pipelines for these science goals (which *could* serve to generate photo-z for the `Object` catalog, as discussed in § 2.5).

8.2.2 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.

8.2.3 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 (SFR). 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 (especially if the SED templates are associated with intrinsic galaxy properties like mass, metallicity, or star formation rate). 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).

8.2.4 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.

8.2.5 Clustering

Photometric redshifts would likely be used by individuals studying large scale structure and galaxy clustering – for example, as a way to make an initial selection of cluster members.

8.2.6 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.

8.2.7 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.

8.2.8 Science Use-Cases Summary

Aside from the specialized use-cases related to dark energy cosmology, which will be served by customized photo-z estimators developed within DESC, most other scientific scenarios use the `Object` photo-z as point estimates of distance in order to subset the data and identifying targets of interest for follow-up, and/or infer intrinsic galaxy properties.

8.3 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.

9 LSST Documentation Review

This section contains a review of all appearances of the terms "photometric redshift", "redshift", or "photo-z" in the LSST documentation. The purpose of this review is to clarify the scope and expected deliverable quality of the `Object` catalog photo-z, and any internal use-cases (as discussed in further detail in § 8.1).

9.1 Science Requirements Document

One of the main science drivers of the LSST design is a significant advance in constraining the models of dark energy cosmology. Section 2.1 of LPM-17 describes the statistical accuracy of photo-z estimates for $i < 25$, $0.3 < z_{\rm phot} < 3.0$ galaxies which are required for the cosmological probes: root-mean-square error $< 0.02(1 + z_{\rm phot})$, bias < 0.003, and fraction of catastrophic outliers $< 10\%$. The SRD specifies that these target statistical values *"are the primary drivers for the photometric depth of the main LSST survey."* In other words, the LSST 10-year photometry must enable a state-of-the-art photometric redshift estimator to achieve these targets – they do not apply to the general-use `Object` catalog photo-z which are the topic of this document.

9.2 Observatory System Specifications

There is a requirement that *"the object catalog completeness"* shall be determined by the DMS for *"a variety of astrophysical objects"*, which includes *"small galaxies on both the red- and blue-sequence at a range of redshifts, and supernovae at a range of redshifts"* [LSE-30]. For the DMS to meet this requirement and determine the object catalog completeness for these classes of objects, it requires redshift estimates for `Objects`. Although spectroscopic redshifts could be obtained and used for this purpose, that would require observing time with non-LSST facilities, and it is instead more feasible that photometric redshifts would be used to meet this requirement. OSS-REQ-0164

9.3 Data Management System Requirements

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]. DMS-REQ-0046

No discussion or details are provided regarding how or when the `Object` photo-z are to be

calculated, validated, or served, or whether it might be equivalent to serve photo-z computed by a third party (i.e., to federate a user-generated photo-z catalog). It is the current role of this document to evaluate the options for fulfilling this requirement and initiate an LSST Change Request to clarify the computation of `Object` photo-z.

9.4 Data Products Definitions Document

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].

9.5 Data Management Science Pipelines Design

This document clarifies that the photo-z estimator would not be developed by LSST DM, but that DM would be responsible for implementing the code to run on the entire `Objects` catalog and validating the results:
"In addition to data products produced by DM, a data release production also includes official products (essentially additional Object table columns) produced by the community. These include photometric redshifts and dust reddening maps. While DM's mandate does not extend to developing algorithms or code for these quantities, its responsibilities may include validation and running user code at scale" [LDM-151].

References

[1] Beck, R., Lin, C.A., Ishida, E.E.O., et al., 2017, MNRAS, 468, 4323 (arXiv:1701.08748), doi:10.1093/mnras/stx687, ADS Link

[2] Carrasco Kind, M., Brunner, R., 2013, TPZ: Trees for Photo-Z, Astrophysics Source Code Library (ascl:1304.011), ADS Link

[3] Carrasco Kind, M., Brunner, R.J., 2014, MNRAS, 441, 3550 (arXiv:1404.6442), doi:10.1093/mnras/stu827, ADS Link

[4] **[LSE-79]**, Claver, C., The LSST Commissioning Planning Team, 2017, *System AI&T and Commissioning Plan*, LSE-79, URL https://ls.st/LSE-79

[5] **[LSE-30]**, Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2018, *Observatory System Specifications (OSS)*, LSE-30, URL https://ls.st/LSE-30

[6] Crocce, M., Ross, A.J., Sevilla-Noarbe, I., et al., 2019, MNRAS, 482, 2807 (arXiv:1712.06211), doi:10.1093/mnras/sty2522, 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] Graham, M.L., Connolly, A.J., Ivezić, Ž., et al., 2018, AJ, 155, 1 (arXiv:1706.09507), doi:10.3847/1538-3881/aa99d4, ADS Link

[10] Gschwend, J., Rossel, A.C., Ogando, R.L.C., et al., 2018, Astronomy and Computing, 25, 58 (arXiv:1708.05643), doi:10.1016/j.ascom.2018.08.008, ADS Link

[11] Hildebrandt, H., Arnouts, S., Capak, P., et al., 2010, A&A, 523, A31 (arXiv:1008.0658), doi:10.1051/0004-6361/201014885, 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] 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

[15] Polsterer, K.L., D'Isanto, A., Gieseke, F., 2016, arXiv e-prints, arXiv:1608.08016 (arXiv:1608.08016), ADS Link

[16] Sánchez, C., Bernstein, G.M., 2019, MNRAS, 483, 2801 (arXiv:1807.11873), doi:10.1093/mnras/sty3222, ADS Link

[17] Sánchez, C., Carrasco Kind, M., Lin, H., et al., 2014, MNRAS, 445, 1482 (arXiv:1406.4407), doi:10.1093/mnras/stu1836, ADS Link

[18] **[LDM-151]**, Swinbank, J.D., et al., 2017, *Data Management Science Pipelines Design*, LDM-151, URL https://ls.st/LDM-151

[19] Tanaka, M., Coupon, J., Hsieh, B.C., et al., 2018, PASJ, 70, S9 (arXiv:1704.05988), doi:10.1093/pasj/psx077, ADS Link

[20] The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al., 2018, arXiv e-prints, arXiv:1809.01669 (arXiv:1809.01669), ADS Link